Lectures sur les mathématiques, l'enseignement & les concours

Volume 1



L. Chélamie, D. Hoareau, R. Rolland,
J.-E. Rombaldi, A. de Saint Julien
Directeur de publication: Dany-Jack Mercier

Lectures sur les Mathématiques, l'Enseignement & les Concours

Volume I

Textes rassemblés et présentés
par Dany-Jack Mercier

L. Chélamie — *L'apprentissage de l'autonomie devant un énoncé mathématique en classe de sixième.*
D. Hoareau — *Histoires de groupes.*
D.-J. Mercier — *Introduction aux espaces projectifs, preuves des théorèmes de Pappus et de Desargues, dualité.*
D.-J. Mercier — *Polyèdres eulériens et solides pathologiques.*
R. Rolland — *Outils élémentaires de l'analyse.*
J.-E. Rombaldi — *Accélération de la convergence des suites réelles.*
A. de Saint Julien — *Matrices toutes puissantes.*

Table des matières

Préface

C'est avec beaucoup d'intérêt que j'ai pris connaissance du projet de Dany-Jack Mercier d'éditer une revue originale destinée à publier des travaux, consacrés aux mathématiques et à leur enseignement d'un type dont on doit bien admettre qu'il a du mal à trouver sa place dans les revues existantes.

Fruits de l'expérience de leurs auteurs, comme enseignants dans le secondaire, dans le supérieur, et notamment en préparation aux concours de recrutement d'enseignants, également comme chercheurs, ces articles possèdent un caractère hybride, métissé, qui les fait échapper aux canons classiques de l'article de didactique, historique, épistémologique, mathématique ou du manuel de cours.

Et pourtant, c'est bien la lecture de ce type de contributions, érudites sans être académiques, rigoureuses sans être pesantes, pleines de remarques, de conseils ou de compléments en marge du propos principal, qui aide à fonder une culture mathématique riche et cohérente, à l'opposé de la succession d'éléments épars et sans lien qui marque trop souvent les cursus actuels.

Connaissant l'exigence de rigueur mathématique de Dany-Jack Mercier, rigueur qui traverse son oeuvre didactique, scientifique au travers de projets aussi divers que réussis - comme son site Mégamaths, ses ouvrages pédagogiques - je peux affirmer que le lecteur pourra parcourir les articles en toute confiance et qu'il y fera une multitude de découvertes enrichissantes.

Dans une période où l'immédiat, le clinquant, la superficialité semblent l'emporter, il est rassurant de voir quelqu'un se lancer seul, dans une entreprise aussi noble qu'utile à la collectivité. Longue vie aux "Lectures sur les Mathématiques, l'Enseignement & les Concours" !

Antoine Delcroix
Professeur à l'IUFM de Guadeloupe
Responsable du CRREF
Président du Conseil Scientifique et Pédagogique

Avant-propos

Voici le premier numéro d'une revue dédiée aux mathématiques, à l'enseignement et aux concours.

Vaste programme s'il en est, cet espace de liberté est ouvert à tous les collègues de la maternelle à l'université qui désirent partager leurs travaux pour donner à ceux-ci une visibilité supplémentaire et les proposer aux lecteurs avertis. Les thèmes abordés dans cette revue se veulent libres et variés.

En me lançant dans ce projet de publication, ma première idée était de donner la parole aux praticiens des mathématiques, à tous mes collègues qui vivent leur science au jour le jour devant des élèves, en leur proposant un espace où ils pourraient s'exprimer en direction de tous ceux qui désirent lire de beaux textes mathématiques, qui veulent réfléchir sur les divers aspects de notre métier d'enseignant, ou qui préparent des concours et cherchent aides et munitions sous la forme d'exposés ciblés et suffisamment détaillés sur des thèmes fondamentaux.

Mon expérience de webmestre de MégaMaths m'a montré qu'un bon nombre de "perles", de travaux intéressants faits par des collègues, ne sont pas diffusés largement ou, s'ils le sont sur internet, font rarement l'objet d'une publication officielle. C'est dommage, car même si la mise en ligne sur internet est un outil remarquable et simple de mise à disposition de l'information, la publication traditionnelle, avec ses besoins de présentation minutieuse et de relectures répétées, demeure le moment où la publication "prend date" et se trouve définitivement prise en compte par la communauté.
Aller jusqu'à la publication physique d'un livre, c'est aussi proposer aux lecteurs un outil de travail bien réel qu'il pourra avoir plus envie d'utiliser.

Cet espace est un espace de liberté où nous nous retrouvons entre "aficionados" : auteurs et lecteurs. Ici seul le contenu et les idées importent, sans qu'il soit nécessaire de suivre une ligne imposée ou de se plier à une quelconque mode forcément passagère.

Ce premier numéro propose 7 articles. Le premier est un article de didactique qui montre le travail d'un professeur dans sa classe de sixième, et les six

[0] revue1 v2.00

autres sont des articles de fond qui intéresseront en particulier les candidats aux concours qui désirent réviser un thème et s'entraîner sur des exercices corrigés.

Voyons cela plus en détail :

▶ Dans "*L'apprentissage de l'autonomie en sixième*", Laurence Chélamie nous montre, exemples à l'appui, comment un professeur de collège peut espérer rendre ses élèves plus autonomes et directifs devant un problème mathématique. L'objectif est bien de motiver un authentique "passage à l'acte" pour qu'un élève de sixième ne se sente plus démuni devant un texte mathématique qui lui est proposé.
Comment lui donner des schémas de réaction ? Quels objectifs doit-on choisir ? Quelles activités peut-on imaginer dans sa classe pour atteindre ces objectifs ? Autant de questions qui méritent d'être posées et qui le seront, de façon toujours très sensible, dans cet article.

▶ Ensuite nous laisserons notre collègue et pédagogue Dominique Hoareau nous raconter des "*Histoires de groupes*". Son article de synthèse permet de survoler en quelques pages une bonne partie de tout ce que l'on doit connaître sur les groupes quand on prépare un concours de recrutement.
Les structures algébriques ont été une des conquêtes du vingtième siècle, et les groupes tiennent une place de choix dans ces structures. L'exposé est bien mené et motivant !

▶ Je vous propose ensuite de plonger dans un espace projectif pour aller voir comment démontrer d'un seul coup deux jolis et très classiques théorèmes d'alignement, les théorèmes de Pappus et de Desargues, qui n'ont rien perdu de leur fraîcheur malgré les siècles !
Démontrer ces résultats sans devoir nous placer dans chacun des six cas de figures possibles est une bénédiction.
Mais pour ce faire, il est nécessaire de bien introduire les espaces projectifs et de montrer en quoi ils peuvent nous intéresser en géométrie classique. J'ai donc insisté sur la description du lien entre espaces projectifs et espaces affines, et sur toute la mécanique du travail "en coordonnées projectives" une fois que l'on a choisi un repère projectif.
Aucune connaissance préalable des espaces projectifs n'est demandée. Il suffit de connaître un peu d'algèbre linéaire pour pouvoir lire cet article. La fin de l'exposé profite de l'investissement que nous avons consenti pour parler :

- des théorèmes duaux (qui se déduisent mécaniquement d'autres théorèmes en échangeant les rôles des points et des droites),

- des homographies et du lien avec les fonctions homographiques de $\mathbb{R}$ dans $\mathbb{R}$.

▶ Le quatrième voyage se passe au pays des solides pathologiques et donc de la célèbre formule d'Euler $S - A + F = 2$ conjecturée en 1750 et démontrée rigoureusement en 1794 par Legendre. La preuve de cette formule sera donnée en utilisant des graphes connexes.

▶ L'article de Robert Rolland est un article de fond sur l'analyse, ses outils de base et leurs utilisations dans le cadre du calcul infinitésimal.
L'auteur sait magnifiquement nous intéresser tout au long de cette pérégrination : accroissements finis, points fixes, méthode de Newton, intégration, interpolation... Le panorama qui se dégage nous permet de mieux prendre conscience des enjeux et des objectifs de cette partie des mathématiques.

▶ Jean-Etienne Rombaldi nous propose ensuite une étude sur "*l'accélération de la convergence des suites réelles*".
Toutes les définitions sont clairement précisées (convergence lente, géométrique de rapport λ, rapide, super-linéaire) et commentées. Pas moins de 19 exercices sont livrés avec une correction complète et serviront d'entraînement pour les candidats aux concours.
L'article, qui s'achève sur la description des procédés d'accélération d'Aitken et de Richardson, a le mérite de proposer un exposé cohérent et précis sur un sujet qui n'est pas souvent traité dans la littérature bien que présent dans le programme de l'oral du CAPES (session de 2008) et donc aussi dans celui de l'agrégation.

▶ Et nous arrivons à la dernière étude proposée par Arnaud de Saint Julien. Notre collègue, enseignant en CPGE, se propose de déterminer toutes les matrices carrées "toutes puissantes" sur un corps $\mathbb{K}$ lorsque $\mathbb{K}$ est $\mathbb{C}$, $\mathbb{R}$, $\mathbb{Q}$ ou un corps fini $\mathbb{F}_q$, c'est-à-dire toutes les matrices carrées A telles que, pour tout $n \in \mathbb{N}^*$, il existe une matrice B vérifiant $A = B^n$.
Cette étude spécialement bien menée est l'occasion de reparler de la réduction des endomorphismes et d'utiliser la décomposition de Dunford.
Beaucoup de jolis résultats sont mis en valeur et utilisés, comme ce raffinement de la surjectivité de l'exponentielle complexe qui s'énonce :

> "Pour toute matrice M de $GL_p(\mathbb{C})$, il existe un polynôme P de $\mathbb{C}[X]$ tel que $M = \exp(P(M))$"

dont on propose deux preuves totalement différentes, l'une en utilisant la décomposition de Dunford et le Théorème de Cayley-Hamilton (Lemme 7.2) et l'autre, très élégante, qui utilise des arguments topologiques (exercice 7.3).
Cerise sur le gâteau, la dernière section propose quatre exercices originaux et bien jolis qui mettent en oeuvre des résultats importants.

Voilà qui clôt ce petit tour d'horizon du volume I.

Je voudrais terminer en remerciant chaleureusement tous les contributeurs de ce numéro, et je dois dire que, malgré le temps et la sueur que cela m'a coûté, j'ai pris suffisamment de plaisir à lire ces articles tout en les formatant pour qu'ils entrent dans ce recueil aux normes imposées... pour recommencer l'aventure prochainement pour un second volume.

Je voudrais aussi signaler que j'ai entrepris ce travail dans le cadre du CRREF (Centre de Recherche et Ressources en Education et Formation de l'IUFM de Guadeloupe) auquel je fais partie.

Je lance enfin un appel aux collègues qui voudraient me proposer un article en leur demandant de prendre simplement contact avec moi par mail[1]. Bien entendu vous pouvez aussi me contacter pour me faire part de vos remarques au sujet du volume que vous avez entre les mains. Vous pouvez aussi aller sur le site Web MégaMaths pour me retrouver, partager vos commentaires, télécharger des documents numériques ou découvrir l'espace qui sera dédié à cette revue.

Maintenant il est temps de nous plonger dans ces articles passionnants... et citer Charles Baudelaire :

Plonger au fond du gouffre,
Enfer ou ciel, qu'importe ?
Au fond de l'inconnu
Pour trouver du nouveau !

Bonne lecture !

Pointe-à-Pitre ce 26 janvier 2009
Dany-Jack Mercier

AVERTISSEMENT :
Ce livre a fait l'objet d'une première parution aux éd. Publibook en 2009
Photo de couverture extraite du catalogue de Gwenn Seemel :
Bohemian waxwing, sur Flickr

[1] dany-jack.mercier@hotmail.fr

Chapitre 1

L'apprentissage de l'autonomie en sixième

L'apprentissage de l'autonomie devant un énoncé mathématique en classe de sixième.

(Laurence Chélamie [1])

Résumé : En entrant en sixième, l'enfant découvre un nouveau contrat avec les professeurs, doit acquérir de nouvelles habitudes et conquérir plus d'autonomie. Mais comment permettre aux élèves d'être plus autonomes et directifs devant un problème mathématique ? Quelles activités peut-on construire pour faciliter ce « passage à l'acte » nécessaire chez nos élèves ? Que répondre quand ceux-ci se contentent de clamer : « Madame, je ne comprends pas ! » ? Cet article, tiré d'un mémoire professionel présenté en juin 2003 à l'IUFM de Guadeloupe, nous offre des pistes de réflexion sur l'enseignement des mathématiques en classe de sixième, tout en décrivant et analysant des expériences pédagogiques réelles vécues dans une classe[2].

[1] Professeur certifiée au Collège de Port-Louis (Guadeloupe), chelamie.emma@wanadoo.fr.

[2] Je voudrais remercier mon tuteur, M. Dany-Jack Mercier, pour sa disponibilité, son aide et ses conseils. Je tiens également à remercier M. Rony Versin, professeur des écoles, qui m'a guidé dans l'élaboration de ce mémoire. Enfin, j'adresse un vif remerciement à tous les élèves de la classe de sixième du Collège Edmond Bambuck du Gosier (en Guadeloupe) sans qui ce travail n'aurait jamais pu exister.

1.1 Introduction

Au collège, les mathématiques contribuent, avec d'autres disciplines, à entraîner l'élève à la pratique d'une démarche scientifique. L'objectif est alors de développer conjointement et progressivement les capacités d'expérimentation et de raisonnement (observation, analyse, pensée déductive), de stimuler l'imagination et l'intuition, d'habituer l'élève à s'exprimer clairement, aussi bien à l'écrit qu'à l'oral, et de développer son analyse critique.

Pour se faire, on vise la maîtrise des techniques mathématiques élémentaires de traitement d'informations (organisation des données, représentations, modélisations, mises en équation...) et de résolution (calculs, équations, constructions...).

Leur emploi, dans la prévision et l'aide à la décision, est précieux dans de multiples circonstances, allant de la gestion familiale à l'activité professionnelle, et de cette manière les mathématiques contribuent à la formation du futur citoyen, ce qui constitue l'une des missions affichée du collège.

Dans le domaine spécifique de l'enseignement des mathématiques, un certain nombre de conditions sont mises en œuvre pour amener l'élève à développer ses capacités de travail personnel, ainsi que son aptitude à chercher, à communiquer et à justifier ses affirmations.

A ce niveau, sur le terrain, on prend vite conscience des difficultés inhérentes à l'incompréhension d'un énoncé mathématique : cela apparaît en sixième dès les premiers apprentissages, et suscite le plus souvent chez l'élève une perte de confiance en soi ainsi que le désir d'abandonner tout effort pour trouver une réponse au problème posé.

Dans cet article, je propose quelques éléments de réflexion concernant l'apprentissage de l'autonomie d'un élève de sixième face à un énoncé mathématique. C'est l'observation d'élèves de sixième qui m'a incité à réfléchir sur cette « nécessaire » quête de l'autonomie, et à rechercher des situations d'accompagnement qui, autant que possible, prennent en compte l'aspect cognitif, sans pour autant négliger les dimensions affectives et relationnelles de la situation d'apprentissage. Le but de cet accompagnement est d'amener l'élève à se rendre autonome par rapport à un énoncé, c'est-à-dire agir efficacement seul, prendre des initiatives, et cela suppose qu'il ait intériorisé des éléments d'observation, de recherche et de remédiation.

Devant un exercice mathématique, un élève de sixième adopte deux attitudes différentes. On remarque d'un côté des élèves qui cherchent, écrivent et parfois trouvent la solution et, de l'autre, des élèves qui n'écrivent rien, semblent relire l'énoncé à plusieurs reprises, et finissent parfois par interpeller leur professeur

en prétextant « ne pas comprendre » ce qui est demandé, voire « ne pas comprendre un seul mot de l'énoncé ».

Voici une anecdote permettant d'illustrer ces propos. Lors d'une séance de travaux dirigés en géométrie, l'exercice proposé était le suivant :

> Soit un segment $[AB]$ de longueur 6 cm.
> Placer le point O milieu de $[AB]$.

Pendant la phase de recherche individuelle et après entretien dans la classe, j'ai pu constater qu'aucun des élèves ne comprenait la signification du mot « soit ». Pourtant, deux attitudes émergèrent. D'un côté, des élèves, disons le groupe A, qui restèrent focalisés sur le mot « soit », et pour qui il était impossible de poursuivre l'exercice à cause de ce mot. D'un autre côté, le reste de la classe, disons le groupe B, qui occulta le mot « soit », et décida, de façon très logique, que, pour placer le point O, il fallait bien au préalable avoir été capable de tracer le segment $[AB]$.

Avec le groupe A, il fallut expliquer que le mot « soit » voulait simplement dire que l'on considérait un segment $[AB]$, reformulation a priori plus complexe, mais qui paru beaucoup plus claire et permit de débuter la construction. Contrairement aux élèves du groupe A, ceux du groupe B avaient réussi à prendre du recul vis à vis de l'énoncé et ainsi, à faire preuve d'une certaine autonomie face à ce dernier.

Ainsi donc, une simple reformulation de la part du professeur pouvait permettre de débuter la résolution d'un exercice. Cette reformulation orale des énoncés, dès lors que ces derniers étaient incompris par les élèves, étaient certes une réponse possible, mais risquait à la longue de leur porter préjudice, dans le sens où elle ne développait pas leur autonomie mais au contraire les encourageait à rester dépendants de leur enseignant. A long terme, une telle attitude de dépendance vis à vis du professeur risquait d'empêcher l'élève d'acquérir et de maîtriser le vocabulaire spécifique des mathématiques, d'empêcher une prise d'initiative, alors même qu'il serait inévitablement amené à réagir seul face à un énoncé lors des devoirs à la maison ou pendant les contrôles.

Toutes ces raisons m'ont incitée à me poser la question de l'autonomie des élèves devant un énoncé mathématique : comment pourrait-on faire pour encourager la prise d'initiative et le choix d'une attitude autonome, en particulier chez l'élève rencontrant des difficultés en mathématiques ?

Mais que désigne-t-on par l'autonomie d'un élève par rapport à un énoncé mathématique ?

> « Dans le domaine des apprentissages, l'autonomie se caractérise par la faculté de prendre en charge ses apprentissages..., et de mener à bien la tâche demandée » [2].

Certains apprenants ont du mal à mener à bien cette tâche. Dans une classe, les élèves qui ne sont pas autonomes sont vite repérés : ils ont toujours besoin du professeur à leurs côtés, afin que celui-ci seconde chacun de leurs gestes. Ce trait de caractère est relativement fréquent chez l'élève de sixième.

Essayer de rendre l'élève autonome face à un énoncé mathématique, c'est essayer de l'amener à réagir efficacement seul face à ce dernier. Bien entendu, pour qu'il puisse y parvenir, cela sous-entend qu'il ait déjà assimilé quelques éléments de remédiation « que le professeur aura mis en lumière avec lui ».

Conservons en mémoire qu'un des objectifs importants de l'enseignement des mathématiques est d'amener l'élève à pouvoir résoudre des problèmes. Cette résolution passe par différentes phases qui nécessitent à chaque fois des apprentissages spécifiques. Il s'agit en général de :

- comprendre l'énoncé et construire une représentation,
- mathématiser l'énoncé et le mettre en signes,
- mettre en œuvre des stratégies et des procédures de résolution,
- répondre à la question posée et vérifier si cela correspond au « bon sens » ou à l'intuition.

La compréhension de l'énoncé est la première étape à franchir dans le processus de résolution, et c'est souvent à ce niveau que surgissent les premières difficultés. C'est le premier facteur important qu'il faut prendre en compte pour que l'élève devienne un jour autonome face à un énoncé mathématique. Dans cet article, je me limiterai volontairement à ce premier facteur, c'est-à-dire essentiellement à la compréhension de l'énoncé, et en particulier aux problèmes respectifs de la lecture et du décodage de celui-ci.

Ces précisions étant données, il convient dans un premier temps de définir ce que l'on entend par « énoncé » et « consigne », puis d'essayer de mettre en lumière les raisons de l'attitude passive de certains élèves face à un énoncé, pour finalement expérimenter quelques solutions.

1.2 Quelques définitions...

Mais, que signifie un énoncé en mathématique ? Une consigne ?

1.2.1 La consigne

Dans le Petit Larousse, la consigne se définit comme « une instruction formelle donnée à quelqu'un chargé de l'exécuter ».

Selon Jean-Michel Zakhartchouk, par « consigne », on entend « toute injonction donnée à des élèves à l'école pour effectuer une tâche de lecture,

d'écriture, de recherche, etc. La consigne s'appuie souvent sur un énoncé explicite, mais les données nécessaires pour l'effectuer sont parfois implicites, d'où la nécessité d'un décodage (...) » ([11] p. 18).

1.2.2 L'énoncé mathématique

La notion de « problème mathématique », prise au sens large, comprend des exercices, des activités et des problèmes enchaînés. L'énoncé écrit d'un problème mathématique est un texte injonctif, qui se constitue, le plus souvent, de deux parties :

- La partie informative, où l'on donne les informations nécessaires à la réalisation de la tâche. La présentation de ces informations peut se faire sous différentes formes : texte, tableaux, graphiques, dessins, schémas, etc.

- La partie directive, où sont écrites la ou les consignes qui explicitent la tâche à effectuer.

1.2.3 Propriétés de l'énoncé

La consigne, incluse dans l'énoncé mathématique, peut se présenter sous la forme d'une injonction ou d'une question.

Lorsque la consigne est un ordre, la tâche attendue de la part de l'élève est explicite (cf. Exemple 1) et des verbes d'action sont utilisés à l'impératif ou à l'infinitif tels calculer, décrire, tracer...

Lorsque la consigne est une question, l'interrogation peut se présenter sous trois formes différentes : totale, partielle ou indirecte. Dans les trois cas, la tâche attendue de la part de l'élève est alors implicite. En effet, bien que non précisée, une justification est attendue (cf. Exemple 2).

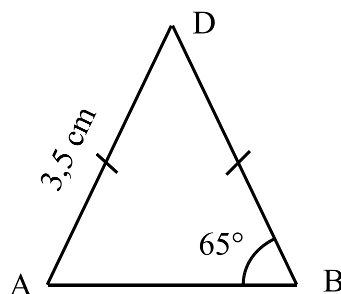


FIG. 1.1 – Figure codée.

Exemple 1 ([7], p. 250) :

▶ *Information* : On donne la figure 1.1 qui est codée.

▶ *Consignes* : Donner la mesure de l'angle $\widehat{BAD}$ et la longueur BD. Expliquer.

Il est intéressant de noter que les énoncés d'exercices de géométrie en sixième, sont plus directifs que ceux d'algèbre ou de gestion de données. En général, ils ne sont composés que de consignes.

Exemple 2 ([4], p. 49) :

▶ *Informations* : Dans une salle de cinéma, il y a 127 places. Seules trois rangées de 31 sièges sont occupées.

▶ *Consigne* : Combien y a-t-il de places libres ?

1.3 Difficultés rencontrées par les élèves

1.3.1 « Madame, je ne comprends pas ! »

Pour trouver les raisons de l'attitude passive de certains élèves face à un énoncé, je me suis focalisée sur cette phrase qui ressort dès qu'on les interroge :

« Madame, je ne comprends pas ! ».

Que peut-il se cacher derrière cette phrase ?

Après quelques questions, je me suis rendue compte que le champ lexical d'un élève de niveau moyen était limité. Cette même phrase avait une signification différente pour chacun des élèves. Les sous-entendus les plus récurrents de cette phrase sont les suivants :

▶ Manque de motivation, découragement et absence de volonté à vouloir s'impliquer dans le travail de recherche demandé. Ainsi, interpeller le professeur par l'intermédiaire de cette phrase, se trouve être, pour eux, la garantie d'une aide.

▶ Incompréhension de l'énoncé ou de la consigne, provenant soit d'une lecture précipitée, soit d'un mal plus profond (ignorance du vocabulaire spécifique et vocabulaire courant, ignorance des constructions et usages syntaxiques, analyse erronée de la phrase. . .).

▶ Définitions non sues, et par conséquent impossible à appliquer.

▶ Incapacité d'organiser les étapes de résolution, on ne sait pas comment s'y prendre.

▶ Incapacité de savoir les notions mathématiques mises en jeu et quels outils mathématiques utiliser.

La motivation est un paramètre important dans le processus d'apprentissage de l'élève et pourrait, à lui seul, faire l'objet d'un mémoire. Dans la suite, nous nous intéresserons seulement aux cas des élèves qui, bien qu'un tant soit peu motivés, restent incapables d'entamer une procédure de résolution.

Afin de mieux cerner les difficultés des élèves, ainsi que leurs méthodes de travail, un questionnaire leur a été distribué (cf. Annexe 1.a et 1.b).

Sur 14 élèves ayant répondu au questionnaire, 9 justifient leur incapacité à débuter un exercice et leur attitude passive, par le fait qu'ils ne comprennent pas la consigne. Selon eux, la solution pour pallier cette difficulté est de relire la consigne jusqu'à ce qu'ils la comprennent. Le cours non appris (43 % des élèves), le manque de temps pour réfléchir (36 %), le découragement et l'énervement (14 %) sont également évoqués.

Concernant la proposition « à partir d'un énoncé de problèmes et de mes connaissances, je sais comment trouver des pistes de solutions », 8 élèves répondent affirmativement, 3 négativement et 3 répondent parfois.

A partir de ce questionnaire, il semble que ces élèves soient capables d'imaginer les causes de leurs difficultés. Malheureusement, leur technique de travail est relativement floue. Mis à part relire l'énoncé lorsqu'ils n'arrivent pas à débuter, il semblerait qu'ils n'aient pas d'autres outils à leur disposition. Ceci corrobore l'idée de Jean-Michel Zakhartchouk suivant laquelle : « Le jeune élève de sixième, bien souvent ne sait pas travailler... » ([11], p. 3).

Quelques difficultés rencontrées par les élèves de sixième face à un énoncé mathématique, semblent devoir être davantage approfondies, afin de pouvoir en apprécier les causes et tenter de trouver des solutions.

1.3.2 Le découragement et la peur d'être jugé

Accepter de résoudre un problème mathématique, c'est accepter de prendre un risque, de se tromper et d'être jugé.

La crainte du jugement peut être un frein dans la prise d'initiatives. Je peux citer comme exemple, le cas d'un élève de ma classe : son caractère est vraiment particulier. Il ne supporte aucune critique et ne s'implique dans les activités que lorsqu'il se sent en confiance. Dans le cas contraire, il ne fournit aucun travail ou vous interpelle en disant qu'il n'a rien compris.

Identifier le type d'erreurs commises lors du devoir en classe :

Je n'ai pas répondu	ex1 (3)						
Je ne savais pas et je ne sais toujours pas							
Je ne savais pas et maintenant je sais							
Je n'ai pas su demarrer							
J'ai perdu du temps : fausses pistes, recherches inutiles							
J'ai perdu du temps en recopiant							
J'ai besoin de plus de temps pour réfléchir							
J'ai donné des réponses fausses							
J'ai mal lu les consignes							
J'ai fait des erreurs de calcul							
J'ai fait des erreurs d'étourderie							
J'ai mélangé plusieurs notions							
Je n'ai pas justifié mes résultats							
J'ai donné une réponse incomplète	X						
Je n'ai pas su utiliser les règles du cours							
Autre explication : (							

1. **Quand vous ne réussissez pas un exercice, quelles sont à votre avis les raisons de votre échec ?**

Soit je n'ai pas appris mes leçons ou je n'ai pas compris la consigne ou le sens de l'exercice.

2. **Lors d'un exercice, vous arrive –t-il de ne pas comprendre ce que l'on vous demande ? Si oui , dîtes les raisons qui font que vous ne comprenez pas ce qui est demandé.**

Oui , souvent c'est que je ne refflechis pas assez et ne li pas bien la consigne ou la question poseē

3. **Que faîtes vous lorsque vous ne comprenez pas un exercice , ou que vous n'arrivez pas à démarrer?**

je relis l'énoncé ou la consigne jusqu'à que je comp

TSVP →

Annexe 1.a

Coche ci-dessous les réflexions qui traduisent ce que tu ressens

	oui	non	parfois
Je trouve que les maths, c'est difficile		X	
Je fais des erreurs bêtes			X
A partir d'un énoncé de problème et de mes connaissances, je sais comment trouver des pistes de solutions	X		
Le cours va trop vite en maths		X	
Je me trompe dans mes raisonnements		X	
Je comprends la correction et je suis capable de la refaire ensuite	X		
Je fais des étourderies			X
Je cherche suffisamment les solutions	X		
Je sais être clair(e) dans mes explications	X		
Je prends le temps de réfléchir	X		
Je comprends le sens des questions posées dans les problèmes	X		
Je sais expliquer mes calculs	X		
Je sais reconnaître les exercices du même genre que ceux qu'on a déjà fait.	X		
Quand j'ai reconnu un exercice du même type qu'un déjà fait, je sais reproduire la solution.	X		
Les devoirs à la maison sont trop nombreux		X	
Je refais souvent les mêmes erreurs en maths		X	
Je sais ce qu'il faut apprendre en maths	X		
J'ai du mal à me concentrer pour chercher		X	
Je travaille sérieusement en maths	X		
Même si ma façon de travailler n'est pas efficace, je n'ai pas envie d'en changer	X		
J'ai peur devant les notions nouvelles		X	
J'ai peur des contrôles de maths			X
Je suis nul en maths		X	

Annexe 1.b

Le découragement est également un frein à cette prise d'initiatives. Certains élèves partent vaincus dès le départ en prétextant que de toutes les façons, ils sont nuls en mathématiques et donc ne peuvent pas comprendre. C'est face à ce type d'attitude que l'on se rend compte que démystifier l'énoncé n'est pas chose inutile.

1.3.3 Les difficultés liées à la langue française

Certains élèves de l'école primaire arrivent en sixième avec une maîtrise imparfaite de la langue française. Par conséquent, ces derniers rencontrent des obstacles au niveau de la lecture et de la compréhension d'un texte, de la gestion des données, de l'argumentation, etc. De plus, « l'apprentissage de la

lecture n'est pas achevé à l'entrée en sixième » ([8], p. 88). L'élève de sixième n'est donc pas capable de lire n'importe quel texte.

La maîtrise imparfaite de la langue, influe sur la compréhension des mathématiques. Les domaines concernés sont le vocabulaire limité, les conjonctions de coordination (alors, puis, et, donc, ...), la syntaxe...

1.3.4 Le vocabulaire

L'une des caractéristiques du langage mathématique est l'usage de mots de la vie courante dans un sens spécifique et ceci, n'est pas sans conséquences. Le sens de certains mots peut être différent selon la discipline dans laquelle il est employé.

Considérons par exemple, le verbe « comparer ». En mathématiques, comparer des nombres signifie dire lequel est le plus petit ou le plus grand, ou encore s'ils sont égaux. En français, il signifie « étudier les ressemblances et les différences entre deux êtres ou deux choses ».

Les différents emplois des mots de la vie courante dans l'enseignement des mathématiques, sont souvent chez eux, source de confusion. Citons comme exemple, les mots « milieu » et « centre ». Dans la vie courante, le mot milieu est employé dans l'expression «être au milieu de ... » et en mathématiques, on parle du milieu d'un segment. Dans le premier cas, « milieu » renvoie à « être au centre de », ou « être parmi... », alors que dans le deuxième cas, il signifie « un point appartenant au segment et équidistant des deux extrémités ».

J'ai encore en mémoire, une anecdote qui s'est déroulée dans la classe de sixième d'un confrère et qui illustre bien la situation. Les élèves devaient réaliser une construction géométrique. La consigne était la suivante : Montrer que I est le milieu de $[AB]$. Un élève l'interpella et lui dit « Tenez monsieur, je vous montre ». Et, il montra du doigt sa figure où, effectivement, le point I était le milieu du segment $[AB]$.

En ayant le souci d'aider à la compréhension de l'énoncé chez l'élève et sachant que « la maîtrise de la langue est un objectif majeur de l'enseignement au collège » ([3], p. 23), il apparaît que définir les mots que nous utilisons en mathématiques est une étape importante, ainsi que donner des exercices où ils doivent employer le mot correct pour décrire une situation (cf. Expérience 2 p. 31).

Nous devons garder en mémoire que la maîtrise de la langue (lecture, écriture et expression orale) est une priorité de l'enseignement au collège et qu'elle « ...ne peut être pleinement assurée que par la contribution de chaque disci-

pline »[3]. La maîtrise du langage à travers l'enseignement des mathématiques, va contribuer à la maîtrise de la langue française.

1.3.5 Les conjonctions de coordination

En interrogeant mes élèves sur leurs cours ou en travaillant avec eux sur des activités, j'ai pu me rendre compte que l'emploi des conjonctions de coordination telles que « si, alors, donc », n'était pas correctement perçu. Très utilisées lors de la rédaction de certaines propriétés mathématiques, elles constituent un obstacle certain à la compréhension de ces propriétés par l'élève. La même difficulté surgit évidemment en géométrie, particulièrement dans les programmes de construction (« construire un triangle ABC tel que ..., puis tracer ... »). Leur emploi dans ces programmes de construction, empêche certains élèves de percevoir l'ordre dans lequel il faut effectuer les étapes. Selon Pierre Legrand [8], cette situation s'explique par le fait que l'élève, entrant en sixième, manque d'entraînement pour pouvoir analyser des phrases à plusieurs verbes, car il n'a pas étudié la phrase complexe.

Tenant compte de cette difficulté, j'ai tendance à ne plus utiliser de conjonctions dans un exercice de géométrie, mais plutôt à numéroter les étapes du programme de construction ou à passer à la ligne à chaque nouvelle étape de la construction. La méthode semble efficace. Mais son emploi systématique ne favorisera pas la maîtrise de ces conjonctions par l'élève.

En se référant aux exercices de géométrie de leur manuel, on se rend compte que très peu de conjonctions de coordination sont utilisées dans les programmes de tracés.

1.4 Le langage mathématique

1.4.1 Un vocabulaire et des notations spécifiques

Enseigner les mathématiques revient en grande partie à enseigner l'utilisation correcte d'un langage spécifique. Il s'agit de faire traduire dans ce langage, qui doit s'efforcer d'être rigoureux, un raisonnement logique qui est plus ou moins confus dans la tête des élèves.

Comme on le dit en [3], l'élève doit être capable d'employer correctement le vocabulaire de l'arithmétique, de la statistique et de la géométrie dans tout type d'activités, et doit pouvoir manipuler les notations spécifiques.

Il faut reconnaître que ce vocabulaire et ces notations sont un obstacle pour l'élève dans son processus de compréhension de l'énoncé.

[3] Extrait d'un Bulletin Officiel de l'Education Nationale.

1.4.2 Un langage spécifique.

Les mathématiques et le français sont étroitement liés de par leur rapport avec le langage. En effet, les notions, les théorèmes, les définitions ne sont accessibles que par divers modes d'expression. On peut citer : la langue naturelle (le français), ainsi qu'une écriture symbolique ayant ses propres règles de fonctionnement, mais qui sont différentes de celles de la langue naturelle.

La langue naturelle utilisée en mathématiques présente des usages grammaticaux peu utilisés dans d'autres contextes. De plus, les tournures spécifiques, de par leur sens ou leur structure, laissent plus d'un élève perplexe. Nous pouvons citer comme exemple : « soit. . . ., on considère . . . ., ABC est un triangle. . . ».

Une compréhension de ces tournures va donc nécessiter un apprentissage qui reposera sur la loi de « l'habitude ». Plus l'élève sera confronté à ce genre de tournures, plus il les maîtrisera.

1.4.3 Complexité de l'énoncé.

Nous devons avouer que les textes mathématiques (en particulier ceux de géométrie) sont souvent complexes. Cette complexité va en grandissant selon les cycles scolaires. Les phrases sont souvent longues et présentent une grande quantité d'informations. Une des raisons de cette complexité est la spécificité des objets mathématiques. En effet, un objet mathématique est souvent déterminé en fonction d'autres objets. Le cercle est déterminé entre autre à l'aide de son centre et un point, la perpendiculaire à une droite par le point dont elle est issue et la droite en question. Ces phrases engendrent également une grande complexité syntaxique des groupes nominaux et font intervenir un grand nombre de prépositions qui y jouent un rôle fondamental. On comprend mieux en quoi la maîtrise de la langue française est importante dans la compréhension des mathématiques.

L'élève doit pouvoir reconnaître et exploiter toutes les informations du texte, afin de résoudre un exercice. Pour beaucoup d'élèves de ma classe, cette tâche est très délicate, voire impossible à réaliser. Cela se justifie sans doute parce que l'élève de sixième n'est pas capable de lire n'importe quel texte (cf. Section 1.3.3).

A en croire un ouvrage du CNDP, « . . . un texte est d'autant plus lisible qu'il contient des indications facilement accessibles quant à la manière de traiter les informations » [1].

Toutefois, selon Jean-Michel Zakhartchouk, l'enseignant ne doit pas pour autant sombrer dans la « dérive du trop facile » ([11], p. 34), attitude qui

consiste à trop guider l'élève à travers l'énoncé, afin d'améliorer la lisibilité et la faisabilité des exercices.

1.4.4 Difficultés liées aux problèmes numériques

Les difficultés rencontrées par les élèves lors de la résolution de problèmes numériques datent de l'école primaire et sont de deux ordres : difficulté à se représenter mentalement les choses et difficulté à les modéliser.

Dans sa conférence donnée à l'IUFM de Guadeloupe en avril 2003, Monsieur Debû affirme que pour pallier les difficultés de modélisation, nous devons développer chez l'élève l'emploi de dessins ou de schémas (patates, schémas segmentaires, ...). Ceci n'étant pas une activité spontanée dans la résolution de problèmes, elle nécessite donc un apprentissage (cf. Section 1.5.4, Expérience 4).

1.5 L'approche expérimentale

Selon Alain Descaves ([5], p. 26), pour améliorer les compétences des élèves concernant la lecture et la compréhension des énoncés, il faut :

- systématiser cette activité,

- varier les énoncés afin de rendre les élèves plus sensibles à leurs caractéristiques,

- développer chez l'élève un apprentissage de la représentation et de la mathématisation,

- rendre l'exercice de résolution d'un problème moins solitaire grâce aux interactions entre élèves.

C'est en nous appuyant sur ces considérations que nous allons bâtir nos expériences qui seront réalisées en classe de sixième.

Parmi les compétences que l'élève doit acquérir pour devenir autonome face à un énoncé mathématique, on peut citer :

- Savoir lire seul l'énoncé silencieusement et jusqu'au bout.

- Etre capable de pointer du doigt ce qu'il ne comprend pas dans l'énoncé, mais ne pas être paralysé par des incompréhensions partielles.

- Pouvoir relier la consigne aux savoirs et savoir-faire acquis précédemment.

Concernant le deuxième point ci-dessus, je me suis efforcée à inciter l'élève à être le plus explicite possible concernant sa difficulté. Je dois avouer que cela est une étape difficile à respecter, dans la mesure où elle nécessite du temps et

demande un effort supplémentaire à l'élève, celui-ci étant amené à expliciter sa difficulté, et non à se cantonner à répondre « je ne comprends rien ». Les réponses au questionnaire témoignent de leur incapacité à véritablement prendre conscience de leurs difficultés. A la question « Savez-vous trouver des pistes de solutions à partir d'un énoncé ? », et sur les 14 élèves ayant répondu au questionnaire, 12 répondirent « oui », ce qui n'est pas forcément vérifié dans la pratique...

1.5.1 Le petit dictionnaire

Comme je l'ai expliqué à la Section 1.3.3, certains élèves ne comprenaient pas la signification de quelques verbes d'action utilisés dans la consigne, et ainsi, n'arrivaient pas à percevoir qu'elle était la tâche qui leur était demandée. Et ce n'était pas en reprenant les explications oralement que la situation se débloquait.

J'ai donc décidé, de définir la plupart des verbes d'action que l'on retrouve dans les consignes et d'en laisser une trace écrite sur leur cahier de cours.

▷ **Expérience 1**

Lors d'une séance de soutien collectif , quelques courts énoncés sont donnés (cf. Annexe 2). L'objectif n'est pas de résoudre les exercices, mais d'essayer d'expliquer la tâche qui est demandée à travers les verbes écrits en gras.

La recherche se fait par groupe de quatre. Après quinze minutes de recherche, les propositions de chaque groupe sont collectées et écrites au tableau. Puis, commence alors un débat pour trouver celles qui traduisent fidèlement le verbe écrit en gras.

Certains verbes comme placer, construire, tracer, effectuer, recopier, mesurer et compléter, ne posèrent pas de problèmes, puisque les élèves y sont confrontés depuis l'école primaire. Le verbe nommer qui est introduit normalement en sixième avec les figures géométriques, ne posa non plus aucune difficulté.

Le débat eut lieu avec les verbes reproduire, justifier, calculer (la longueur de $[AB]$), encadrer et coder. De plus, le terme « consécutif » laissa plus d'un élève perplexe.

Concernant le verbe comparer, seules les quatre meilleures élèves me donnèrent la réponse quasi-correcte. En effet, dans leur définition il manquait le cas où les nombres sont égaux. Mais, les autres élèves, en suivant mes directives, purent compléter la définition.

Le verbe calculer, dans la formulation « calculer la longueur du segment $[AB]$ », fit émerger deux définitions. Pour certains, il fallait mesurer le seg-

ment $[AB]$ et répondre que le segment $[AB]$ mesure... Pour d'autres, il fallait trouver la longueur de ce segment à l'aide de calculs (en particulier l'inégalité triangulaire). Ceux du premier groupe ne faisaient aucune différence entre calculer la longueur du segment $[AB]$ et mesurer le segment $[AB]$. Cette attitude est encore plus suspecte, du fait que dans la formulation « calculer les sommes suivantes », il n'y a aucune ambiguïté et qu'il faut effectuer des opérations.

Dans chacun des exercices suivants, expliquer la tâche qui est demandée à travers les **verbes écrits en gras**. On ne demande pas de résoudre l'exercice.

Exercice 1
Sur la droite graduée ci-dessous, **placer** le point A d'abscisse 2,5.



Exercice 2
Construire 3 points A, B et C non alignés.
Tracer la droite (AB) en rouge.

Exercice 3 :
Effectuer les additions suivantes : 3,27 + 0,1 = ; 71,12 + 13 =

Exercice 4
Recopier et **compléter** les phrases suivantes avec les mots : « addition », « somme », « terme ».
5,35 + 2,7 = 8,05
Cette opération est .... 5,35 et 2,7 sont les ..... 8,05 est .....

Exercice 5
Comment peut-on **nommer** cette droite ?



Exercice 6
Reproduire cette figure sur votre cahier en respectant les dimensions.



Exercice 7
Tracer un segment [AB] de longueur 6 cm. Placer le point O sur ce segment [AB] tel que AO=3 cm. Le point O est-il le milieu du segment [AB] ? (**Justifier** votre réponse).
Coder votre figure convenablement.

Exercice 8
Tracer un segment [EB] tel que EB= 8 cm. Placer un point A sur le segment [EB] tel que AE = 3cm. **Calculer** AB.

Exercice 9
Encadrer les nombres décimaux suivants à l'aide de deux nombres entiers consécutifs : 2,75 ; 3,01 ; 14,87.

Exercice 10
Comparer les nombres entiers suivants : 3 et 5 ; 18 et 18,0001 ; 26 et 12.

Annexe 2

Ce conflit vient peut-être du fait que pour eux, certains verbes comme calculer sont propres à l'algèbre et que d'autres comme mesurer sont relatifs à la géométrie. Car à leur stade d'apprentissage, on ne fait les calculs proprement dits que dans les problèmes numériques et la construction des figures qu'en géométrie.

Finalement, à l'issue du débat, je tranche en disant laquelle des deux propositions est correcte pour le calcul du segment $[AB]$.

Concernant l'expression « encadrer par deux entiers consécutifs », personne ne put répondre. Deux difficultés semblaient surgir de cette phrase : Le verbe encadrer, bien qu'on l'ait déjà défini et l'expression « entiers consécutifs ».

Je pense que, concernant le verbe encadrer, cela doit venir du fait qu'ils n'ont pas encore assimilé cette nouvelle définition. Par contre, pour l'expression « entiers consécutifs », ce doit être à cause d'un manque de vocabulaire. Il me revient ainsi à expliciter cette expression.

Le verbe coder est un terme nouveau introduit en géométrie. Il signifie « mettre le codage mathématique sur une figure présentant quelques particularités ». Les élèves ayant des difficultés ne comprirent pas la phrase « coder-la convenablement ». Ce sont les autres qui expliquèrent cette phrase.

Le verbe justifier, au niveau du sens ne causa pas de véritables difficultés. Ces dernières surgirent lorsque je posai la question suivante : « Si vous rencontrez ce verbe dans un exercice de géométrie, où tireriez-vous vos arguments ? ». Mon objectif était de les sensibiliser à la notion de démonstration.
La majorité des élèves répondirent qu'ils tireraient leurs arguments à partir de ce qu'ils verraient sur leur « dessin ».
A ce niveau de l'année, je leur ai simplement répondu qu'il ne fallait pas toujours tirer leurs arguments de la figure et que l'on travaillerait sur cette question ultérieurement. Je n'ai pas voulu m'étendre davantage, sachant que ce n'était pas l'objet de la séance.

Toutes ces définitions furent collectées sur le cahier de cours et il fut demandé aux élèves de s'y référer dès qu'ils ne sauraient pas quoi faire.

▷ **Constatations**

Après quelques exercices, certains n'ont toujours pas assimilé le vocabulaire et n'ont pas la présence d'esprit de reprendre leur cahier de cours pour s'y référer. Par conséquent, ils continuent à ne pas savoir quelle est la tâche demandée. Ce manque de présence d'esprit témoigne d'un manque d'autonomie qui n'est toujours pas corrigé.

Les autres ont une assimilation progressive du vocabulaire. De temps en temps, certains ont quelques trous de mémoire, mais les autres arrivent à les combler. Oralement nous sommes obligés de redéfinir ces mots régulièrement.

▷ **Avis**

La méthode écrite semble ne pas être très concluante. Elle profite davantage à ceux qui n'ont pas de difficultés et qui ont déjà pris l'habitude d'apprendre. Pour ceux qui en ont le plus besoin, cela semble être un coup d'épée dans l'eau, puisqu'ils ne révisent pas les définitions qui ont été précisées, et ne s'y réfèrent pas non plus. Mais, je pense qu'une majorité d'élèves a perçu l'importance du vocabulaire.

En tout cas, à partir de cet instant, j'ai pris le temps de redéfinir tous les mots spécifiques et en particulier, les verbes, que l'on retrouve souvent dans les exercices.

Le souci de la maîtrise du vocabulaire étant quasi-présente, une autre expérience est menée sur ce thème : celle des figures « téléphonées ».

1.5.2 Les figures « téléphonées »

La maîtrise de la langue est une priorité de l'enseignement au collège. Les mathématiques y contribuent à travers la production de textes mathématiques, aussi bien à l'écrit qu'à l'oral. Chaque élève doit apprendre et parvenir à s'exprimer clairement et c'est à nous, enseignants, de lui faire ressentir la nécessité d'utiliser un langage précis. C'est pourquoi l'ouvrage « Mathématiques en sixième » [3] suggère de travailler avec l'élève sur « l'écriture d'énoncés à destination d'autres élèves », en particulier dans la description de figures « téléphonées ».

L'expérience que je présente ici, avait aussi pour objectifs de :

- montrer à l'élève l'intérêt d'utiliser un vocabulaire précis,
- savoir décrire une figure géométrique,
- savoir suivre des directives,
- forcer l'écoute.

▷ **Expérience 2**

L'expérience se déroule lors d'une séance de deux heures et se présente sous forme de jeu avec une récompense à la clé. La perspective du jeu et de la récompense a pour effet de motiver davantage la classe.

Pour préparer l'activité, je distribue un texte comportant une figure et un énoncé à trous, en demandant de le compléter. Cet énoncé à trous présente un programme de construction de la figure donnée, et permet aux élèves de comprendre ce que l'on attend bientôt d'eux. Cet exercice n'a posé aucune difficulté particulière. Le jeu pouvait donc commencer.

La classe est séparée en trois groupes de cinq élèves. J'ai constitué les groupes de façon à ce qu'il n'y ait pas de grande différence de niveau entre eux. Mis à part cette question d'équilibre entre les différents groupes, je me suis efforcée de mettre au sein d'un même groupe des élèves qui, en temps normal, n'auraient jamais envisagé de travailler ensemble.

La séance se déroule donc de la façon suivante : Chaque groupe choisit un chiffre entre 1 et 6, chiffre qui correspond à une figure géométrique différente. Pendant vingt minutes, chaque groupe travaille sur son programme de tracé. Après ces vingt minutes de recherche, chaque groupe désigne un orateur qui aura pour mission de lire son programme de tracé.

Le jeu se déroule en trois manches. A chaque manche, une équipe expose son programme et un représentant de chacune des deux autres équipes est désigné. Les représentants travaillent simultanément au tableau et doivent réaliser la figure décrite par l'orateur. Tous les élèves, hormis ceux de l'équipe qui exposent doivent réaliser la figure, car si leur représentant au tableau n'y arrive pas, ce sont les figures correctes réalisées dans leur équipe qui leur rapporteront des points.

La répartition des points est la suivante :

• Si la personne au tableau réalise correctement la figure, elle rapporte 6 points à son équipe.

• Chaque figure correcte réalisée au sein d'une équipe rapporte 1 point.

• Si le programme de construction est correct (cela est décidé par le professeur et les élèves), l'équipe qui expose gagne 5 points. Sinon, elle en perd 1.

La personne qui expose n'a, à aucun moment, le droit de donner des explications sur son programme. Elle est uniquement autorisée à répéter ce qu'elle a écrit sur sa feuille.

▷ Les résultats de l'expérience

[On se reportera aux Annexes 3.a, 3.b et 3.c.]

Dans les groupes 1 et 2, la recherche s'est faite individuellement par écrit, puis après une délibération au sein de chaque groupe, un programme de tracé final a été rédigé. Dans le groupe 2, il y a eu une mise en commun des informations, puis un véritable débat avant de se mettre d'accord sur le programme

final. Dans le groupe 3, la recherche s'est faite oralement. Ensuite, ils se sont tous mis d'accord sur le programme qu'ils allaient rédiger. Chacun écrivit sur sa feuille ce qu'il avait entendu.

Le groupe 3 fut le premier à sentir l'importance de définir exactement chaque objet. Je m'explique : ils me demandèrent s'il était encore possible de choisir quelqu'un d'autre pour exposer le programme de tracé, car selon eux, celui qu'ils avaient choisi au préalable ne dirait pas au moment propice « segment », « droite », etc. Par exemple, au lieu de dire « coupe le segment $[AB]$ en E », il dirait « coupe « AB » en E ». Je les autorisai donc à changer d'orateur.

Lorsque Maddly exposa le programme du groupe 2, elle fit bien attention de parler d'angles $\widehat{CBA}$ et $\widehat{CAB}$, mais lorsqu'elle lut les deux dernières phrases de son programme, elle omit le terme « droite » dans « parallèle à (AB) » et « coupe (CB) en L ».
Cette omission ne sembla pas gêner les deux élèves qui étaient au tableau. J'ai donc pénalisé son équipe, ce qui conforta les élèves du groupe 3 dans l'idée qu'il fallait bien définir les objets lorsque l'on s'exprimait.

Puis, vint le tour d'Olivia d'exposer pour le groupe 3. A ce niveau du jeu, son équipe menait largement et normalement, était censée gagner. Tout se déroulait convenablement. Elle fit attention de bien préciser si les objets considérés étaient des droites ou des segments. Les deux élèves au tableau, ainsi que les autres, réalisèrent jusque là, la figure convenablement.
Un léger brouhaha survint lorsqu'elle lut la dernière phrase. En effet, elle oublia de préciser le point par lequel la droite perpendiculaire à la droite (CP) était issue. Ephanielle s'en rendit compte et essaya de la prévenir en le lui soufflant, mais elle n'y porta pas attention. Les élèves au tableau sentirent qu'il manquait une information puisqu'ils ne savaient pas où tracer cette droite. La règle du jeu ne permettant pas aux élèves de poser des questions, mais les autorisant seulement à demander à Olivia de répéter cette dernière phrase, les nombreuses tentatives d'Ephanielle n'attirèrent pas l'attention d'Olivia sur la possibilité qu'il manquait quelque chose dans cette phrase.

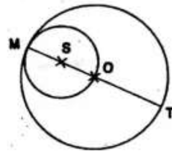
Lorsque je montrai la figure que l'on aurait dû trouver les élèves dirent qu'ils n'avaient pas trouvé la figure correcte, parce qu'Olivia avait oublié de préciser que la droite passait par le point P.

Annexe 3

Résultats du groupe 1

Enoncé

Rédiger un programme de tracé de la figure ci-dessous commençant par « tracer un segment [MT]… »



Les écrits de chaque membre du groupe

Jessy

1) tracer un segment [M T] diamètre de centre o
2) tracer un cercle de centre S et de rayon [MT]
3) tracer le rayon du cercle o placer S
tracer le cercle de centre S passant par [OM]

Sessy

tracer un segment [MT]. O est le milieu de ce segment.
tracer un cercle de centre O tel que MT diamètre de ce cercle
tracer le cercle de diamètre [MO] et de centre S,

Énoncé conservé: celui de Sessy

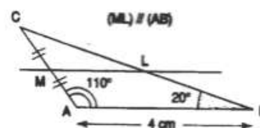
Evodie

tracer un segment [MT] de centre O et placer le compas de façon à ce que [MT] soit le rayon de centre O. Et tracer le cercle MO de centre S.

Annexe 3.a

Résultats du groupe 2

Rédiger un programme de tracé commençant par « tracer un triangle ABC… »



Ecrits des élèves

Harold

tracer un triangle A,B,C de ~~tel~~ que $\widehat{C,A,B} = 110°$, et ~~que~~ $\widehat{C,B,A} = 20$ et que [AB] = 4 cm. placer M le milieu de [C A]. Tracer une droit parallèle à AB passant pour M. La droit coupe [C B] en L.

Maddly

Tracer un triangle ABC, tel que $\widehat{CAB} = 110°$
puis CBA = 20° et que [AB] 4 cm
placer M le milliev de [CA]
tracer un droite paralléle à AB passant par M.
La droite coupe [CB] en L

Programme exposé

Tracer un triangle ABC tel que AB = 4 cm
et $\widehat{CAB} = 110°$ et que $\widehat{CBA} = 20°$.
Placer M, milieu de [CA]
Tracer une droite paralléle à (AB) passant par M
La droite coupe (CB) en L

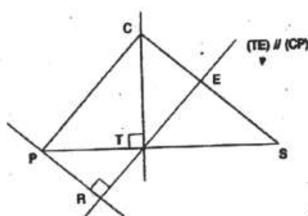
Annexe 3.b

Annexe 3.4

<u>Résultats du groupe 3</u>

<u>Enoncé</u>

Rédiger un programme de tracé de la figure ci-dessous commençant par « tracer un triangle CSP … »



<u>Ecrits des élèves</u>

<u>Joëlle</u>

Tracer le triangle CSP rectangle en C
Tracer une droite perpendiculaire a (PS) passant par C elle coupe le segment [PS] en T. Tracer une parallèles à (CP) passant par T et coupant le segment [ES] en E. Tracer une droite perpendiculaires à (CP) passant par P et coupant (ET) en R

<u>Ephanielle</u>

tracer le triangle SCP rectangle en C
tracer une droite perpendiculair à la droite (PS) passant par C, elle coupe le segment [PS] en T
tracer une droite parallèle à la droite (CP) passant par T et coupant le segment [ES] en E
tracer une droite perpendiculaire à la droite (CP) passant par P et coupent la droite (ET) en R

<u>Olivia</u>

tracer le triangle CSP rectangle en C
tracer un droite perpendiculaire à (PS) passant par C elle coupe le [PS] en T
tracer une droite parallèle à (CP) passant par T et coupant le segment [ES] en E
tracer une droite perpendiculaire à (CP) et coupant (ET) en R

Annexe 3.c

▷ **Avis**

L'expérience fut très enrichissante pour chacun d'entre nous. L'erreur de Maddly, ainsi que celle d'Olivia, me permirent de leur montrer la nécessité de s'exprimer clairement en géométrie. Je pense que tous les élèves se souviendront de cela puisque l'équipe 3 perdit à cause d'une erreur qui pouvait être évitée. Ce fut l'équipe 2 qui remporta la victoire, bien que pénalisée par l'erreur de Maddly.

Tous les programmes avaient été correctement retranscrits oralement et tout le monde, excepté quatre élèves, ont correctement réalisé les différentes figures.

Le programme devant être énoncé oralement, chacun vit la nécessité de se taire lorsque l'orateur s'exprimait. Ils se rappelèrent mutuellement à l'ordre lorsque quelqu'un, différent de l'orateur, avait tendance à bavarder. Le programme de construction énoncé oralement me permit également de travailler sur un énoncé oral avec eux.

Il est intéressant de noter que dans aucun des programmes de tracé, il n'y avait de conjonctions de coordination. Chaque étape était bien segmentée, à l'écrit comme à l'oral. Peut-être que le jeu ne se prêtait pas à l'emploi des conjonctions de coordination, puisque les directives étaient énoncées oralement et que les élèves devaient réaliser la figure simultanément. L'autre explication est qu'ils ont peut-être été influencés par l'exercice qui avait été distribué initialement.

En guise de remédiation, on peut imaginer qu'il soit intéressant de revenir, plus tard, individuellement et par écrit, sur ce même type d'exercices, en imposant cette fois-ci l'emploi des conjonctions de coordination.

1.5.3 La reformulation par les élèves

Au départ, elle était systématique de ma part. J'ai donc changé ma façon de procéder.

L'expérience s'effectue donc durant toute l'année et se présente de la façon suivante : Je ne reformule plus la phrase systématiquement dès qu'un élève dit ne pas comprendre. Je laisse une phase de réflexion ou de consultation.

Très peu d'élèves travaillent en binôme dans cette classe. Rares sont ceux qui demandent de l'aide à leur camarade de table. Les seuls échanges sont synonymes de bavardages.

Pour revenir à l'expérience, si l'élève en est toujours au même point, je soumets son interrogation à une consultation collective. En général, le reste de la classe parvient à reformuler le problème pour l'élève en difficulté. Si ce n'est pas le cas, je peux prendre le relais.

L'avantage de cette méthode est qu'elle permet :

- un détachement de l'élève vis- à -vis du professeur,
- un enrichissement du vocabulaire,
- la vérification de la compréhension des autres élèves qui reformulent.

1.5.4 Emploi d'un brouillon

J'ai constaté en début d'année que peu d'élèves prenaient un brouillon pour réfléchir. Ce dernier n'était utilisé que dans le cas d'activités numériques et ce, uniquement pour effectuer des opérations du type addition, soustraction, multiplication... De plus, à chaque fois, les élèves me demandaient s'ils avaient droit ou non à un brouillon.

Bien qu'ils aient tous un cahier de recherche, il leur paraît inconcevable de pouvoir y faire des ratures. A priori, à leurs yeux, ce cahier doit être conservé propre et ne doit pas faire l'objet de ratures ou d'erreurs. Cette angoisse du «cahier propre» généra un manque d'initiatives concrètes, car les élèves prétendaient réfléchir mentalement.

Ma première mission concernant ce problème fut de systématiser le brouillon et de leur faire prendre conscience qu'ils avaient le droit de se tromper. Il fallait que le brouillon ne serve pas uniquement à faire des calculs, mais devienne un outil de recherche à part entière, en particulier pour les constructions géométriques.

▷ **Expérience 3**

Cette expérience est menée au cours d'une séance de travaux dirigés portant sur les angles. Les deux derniers exercices de la séance portent sur la construction d'un triangle, avec au préalable la connaissance d'un angle et de la mesure de la longueur de deux côtés, ou encore de deux angles et de la mesure de la longueur d'un côté.

Ce type d'exercices est une première pour eux. Certes, la construction de triangles, à partir de la connaissance de la mesure des longueurs des trois côtés, part du même principe, mais ici, la donnée des angles a tendance à compliquer la tâche. Les meilleurs élèves s'en sortent relativement sans difficultés, ce qui n'est pas le cas du reste de la classe. Ces derniers ne savent quelle démarche adopter et peinent énormément à traduire l'information concernant les angles.

A ce niveau de l'année, l'emploi du rapporteur (nouvel outil introduit en classe de sixième) ne pose pas de problème véritable. La difficulté surgit dans la recherche de la démarche à adopter. Les bons élèves sont sollicités pour

donner une méthode. Mais, il leur est impossible d'en donner une, sans pour autant révéler le processus de construction.

J'introduis donc l'idée de faire une figure à main levée. La démarche est nouvelle pour eux. Pour beaucoup d'entre eux, réaliser une figure géométrique est une tâche qui s'opère uniquement à l'aide d'instruments.

Je dessine donc un triangle quelconque au tableau et les invite à le reproduire sur leur cahier. Ensuite, je leur demande de porter les informations du texte sur la figure, à savoir les noms des sommets, la longueur des côtés et la mesure des angles. Les difficultés apparaissent au moment de placer les angles sur la figure ; ce qui révèle que certains ne maîtrisent pas encore les notations mathématiques des angles.

A partir de la figure à main levée, qui sert de support visuel, je demande de construire le triangle proposé dans le premier exercice.

Cinq élèves sur dix-neuf ne surent dans quel ordre il fallait employer les instruments. Après avoir élucidé ce problème (en partant de ce que l'on sait), ils furent en mesure de construire le triangle.

▷ **Constatations**

Mis à part les erreurs venant de la mauvaise position des angles, certains élèves furent conditionnés par la figure à main levée à tel point qu'ils refusèrent de valider la figure correcte qu'ils avaient obtenue, en prétextant qu'elle ne ressemblait pas à la figure dessinée sur le brouillon. Ils oublièrent que ce brouillon n'était là que pour aider à la construction, et qu'en aucun cas, le triangle recherché était similaire à la figure obtenue sur le brouillon.

Outre ce fait, la méthode fut adoptée pour la construction du deuxième triangle et permit à plusieurs d'entre eux de réaliser la construction sans mon aide. Le fait d'avoir introduit l'emploi de la figure à main levée sur un type d'exercices qu'ils n'avaient pas encore rencontré, m'a permis de leur montrer que le dessin à main levée était un véritable outil de recherche.

Ainsi, l'emploi du brouillon pour aider à la construction de figures géométriques fut systématisé.

Le point positif de cette expérience est que cette démarche soit facilement adoptée par la majorité des élèves, leur offrant un moyen de réaliser des constructions géométriques, tels que des triangles, des losanges etc.

Une deuxième expérience menée sur l'emploi du brouillon et sur le dessin d'une figure à main levée, est décrite dans le paragraphe suivant. Elle a pour objectif de montrer que l'on peut parfois utiliser une figure à main levée, pour donner du sens à un problème numérique, et par conséquent s'aider à raisonner.

▷ **Expérience 4**

Cette expérience est menée lors d'une séance de correction d'un devoir à la maison (cf. Annexe 4). Ce devoir fut une véritable hécatombe et en particulier, pour le deuxième problème. Treize élèves n'eurent pas la moyenne et parmi eux, neuf élèves eurent une note inférieure à 07. Cinq élèves ne réussirent pas l'exercice 1, le plus souvent, à cause d'une mauvaise mise en signes du problème. Au lieu d'effectuer une soustraction, ils effectuèrent une addition.

DEVOIR A LA MAISON N° 3
(A remettre le Samedi 14 Décembre 2002)

Problème de taille

Debout sur un tabouret de 50 cm de haut, Eve « mesure » 1,96 m.
Guy, debout sur une table de 75 cm de haut , « mesure » 2,17 m.
Léa , debout sur une chaise de 4,3 dm « mesure » 1,87 m.
Ranger ces enfants, par ordre croissant, selon leur taille réelle.

Problème

Des agriculteurs ont placé du fil barbelé autour d'un terrain. Ils ont payé ce fil 76€ . Le fil barbelé coûte 4 € le mètre.
Le terrain est un triangle RIZ isocèle en Z. La base de ce triangle mesure 400 cm.
Combien mesurent les côtés [ZR] et [ZI] ?

Annexe 4

Cette erreur aurait pu être évitée à l'aide d'un dessin (FIG. 1.2). Je l'introduisis pour le premier cas et leur demandai de porter les dimensions au niveau des flèches. Ensuite, je les invitai à faire de même pour les autres cas, puis de répondre au problème.

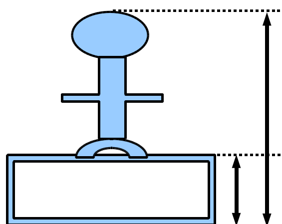


FIG. 1.2 – Un dessin salvateur !

La plupart de ceux qui s'étaient trompés surent se corriger et expliquer leur démarche.

Le deuxième problème présenta plus de difficultés. Seuls huit élèves sont parvenus à le résoudre, dans un premier temps. En effet, pour onze d'entre eux, la somme des mesures des côtés de même longueur était 400 cm, soit $ZR = ZI = 200$ cm. Là encore, je leur proposai de faire un dessin à main levée afin de résoudre le problème. Le manque de support visuel leur a fait défaut. La recherche de la solution prit un certain temps, mais au moins à partir du dessin ils arrivèrent à proposer des idées de solution.

1.5.5 Localisation des « hypothèses » et de la « tâche »

Savoir repérer les informations d'un énoncé et les tâches à effectuer, est une compétence à acquérir afin d'envisager sereinement la résolution de problèmes. Dans les exercices canoniques des manuels, cette faculté n'est pas nécessaire. Ce qui est tout autre dans la résolution de véritables problèmes numériques ou de géométrie. En général, la forme de l'énoncé (cf. Section 1.2.2) peut faciliter le travail. L'expérience suivante va consister à repérer les deux parties de l'énoncé : la partie informative et la partie directive.

▷ **Expérience 5**

Cette première expérience consista à travailler sur des énoncés de problèmes numériques. Il fallait localiser la partie informative, en tirer toutes les informations, puis dire quelle était la tâche demandée ; ce qui revenait à chercher la consigne.

Cette recherche ne posa de soucis à personne. La forme générale des énoncés numériques facilita énormément leur travail.

▷ **Expérience 6**

Cette deuxième expérience se déroula en séance de soutien et porta cette fois sur un énoncé de géométrie. Il s'agissait d'y rechercher les informations, afin de tenter de faire une démonstration. Les élèves qui sont en soutien rencontrent énormément de difficultés pour prouver un résultat. Souvent, ce genre de question est délaissé.

Parmi les raisons supposées, nous pouvons évoquer d'une part, l'incompréhension de ce qui est précisément attendu à travers les verbes « démontrer », « prouver » et « montrer », et d'autre part, la méconnaissance des arguments à avancer pour prouver ces résultats. Bien que la démonstration ne soit réel-

lement traitée qu'en classe de quatrième, les textes officiels nous préconisent, en sixième, à initier l'élève au raisonnement déductif.

Afin d'aider l'élève dans sa recherche, une feuille méthodologique est distribuée (cf. Annexe 5). Personne ne rencontra de difficultés pour répondre à la question B. La recherche de la réponse de la question A me permit de me rendre compte que 5 élèves sur 9 ne savaient pas repérer les informations de ce texte.

En effet, selon ces cinq élèves, les informations disponibles dans le texte étaient que le triangle TIC est rectangle en I, que le point I est le milieu du segment $[TA]$ et que la droite (CI) est la médiatrice du segment $[TA]$.

Dans ce cas de figure, ces cinq élèves assimilent ce que l'on doit démontrer à une information donnée. Pierre Legrand, dans « Profession Enseignant » [8], explique que cette erreur part du fait que, dans certains énoncés, la consigne soit souvent mêlée à la partie informative. Une telle position de la consigne rend aussi bien difficile la distinction entre les deux parties (informative et directive), que la lecture d'énoncés en elle-même.

Des expériences similaires ont été renouvelées à travers d'autres exercices de géométrie et les mêmes difficultés sont réapparues.

Soutien : le 21/02/03

Problème : Soit TIC un triangle rectangle en I.

1. Construire le point A tel que I soit le milieu de [TA].
2. Montrer que la droite (CI) est la médiatrice du segment [TA].

Méthodologie

A) Faire la figure et la coder convenablement.

B) Que veut –on démontrer ?

C) Que nous faut –il pour arriver à ce que l'on veut démontrer ?

D) Qu'avons nous comme données /informations dans l'énoncé ?

E) Exploiter les données/informations de l'énoncé pour arriver à ce que l'on veut (étape C).

Annexe 5

La consigne mêlée à la partie informative empêche certains élèves de la classe, en particulier ceux qui rencontrent des difficultés en mathématiques, de repérer les informations essentielles d'un énoncé mathématique.

La méthode, qui consiste d'abord à repérer la consigne (étape simple, puisque c'est la phrase où l'on demande d'effectuer une tâche), puis de dire que ce qui reste est la partie informative, ne semble efficace que pour les problèmes numériques.

Face à des énoncés de problèmes de géométrie, toutes les compétences relatives à la compréhension d'un texte sont nécessaires, car elles permettent de repérer les informations essentielles d'un énoncé.

1.5.6 Fiches méthodologiques.

▷ **Expérience 7**

L'expérience de la fiche méthodologique a été menée lors d'une séance de soutien et a déjà été présentée dans le paragraphe précédent (cf. Annexe 5). L'objectif était d'introduire une démarche, pour que les élèves puissent entamer un raisonnement déductif et par conséquent, le démystifier.

Le problème traité est extrait d'un devoir à la maison. Aucun des élèves présents à cette séance de soutien n'a répondu à la deuxième question. La recherche fut très laborieuse, en particulier aux questions C et D.

▷ **Constatations et remédiation**

Bien que les caractéristiques de la médiatrice soient connues, l'incapacité d'exploiter les informations a quelque peu terni la portée de cette fiche. Les élèves n'ont pas perçu, dans un premier temps, que la fiche méthodologique avait pour objectif de les aider dans leur démarche.

Il aurait peut-être été plus approprié de l'utiliser sur des exemples plus simples et plus nombreux. Ainsi, le caractère répétitif de l'acte leur permettrait de prendre conscience des différentes étapes décrites sur la fiche.

Cette situation d'échec m'a permis de reconsidérer l'emploi de la fiche méthodologique. Bien qu'ici, elle n'ait pas eu la portée souhaitée, je ne renie pas pour autant son efficacité, car elle peut permettre à certains élèves de débuter la recherche d'un exercice.

Il me reviendra, dans l'optique d'une meilleure approche en géométrie, de consacrer une séquence, pour aboutir à la réalisation d'une fiche définissant « le contrat méthodologique et pédagogique » en relation avec l'énoncé.

Toutefois, je réitérai l'expérience de la fiche méthodologique, mais cette fois, dans le cas d'un problème numérique. L'expérience se déroula dans le cadre

d'une séance de soutien, où la fiche méthodologique (cf. Annexe 6) proposée a été tirée du manuel Dimathème 6e. L'exercice problème considéré avait été donné préalablement lors d'un contrôle.

La partie « repérer ce qui est important en résumant les données et les questions », se déroula sans heurts, en raison du travail déjà effectué sur ce genre de démarche. La difficulté survint lors du résumé des données, sachant que pour cette première fois, c'est moi qui notai, de manière concise, les données au tableau.

La situation est imaginée et retranscrite oralement par les élèves. Ils décomposent chaque étape de la situation en la reformulant avec leurs propres mots. Afin de les aider dans leur recherche, je leur pose quelques questions intermédiaires. Par exemple, « Que cherchons-nous ? Que nous faut-il pour connaître le prix du panier de basket ? A quoi correspond le prix total ? Que nous faut-il calculer pour connaître le prix total ? ».

Enoncé

A la fête de l'école ,les enfants ont vendu 84 gaufres et 5,2 litres de limonade. Une gaufre coûte 1,15 € et le litre de limonade coûte 2,15 € .Avec la somme récoltée, la coopérative de l'école a pu acheter un panneau de basket. Après l'achat de ce panneau de basket, il ne restait plus que 25,6 € pour le voyage de l'école.
Quel était le prix du panneau de basket ? (Toutes les opérations seront posées)

1)Lecture et début de recherche

Repérer ce qui est important dans le texte en résumant les données et les questions.

Ce que je sais	Ce que je cherche

Imaginer la situation.

Peut-être se poser des questions intermédiaires.

2)Solution
Remarque : Chaque opération est suivie de sa phrase-réponse.

Annexe 6

Suite aux réponses apportées, chacune des opérations est alors écrite, suivie d'une phrase-réponse.

La méthode exposée sur cette fiche méthode a été répétée oralement à chaque correction de problèmes numériques, surtout lorsque certains d'entre eux rencontraient des difficultés.

L'idéal serait qu'ils parviennent, systématiquement, à se poser ce genre de questions. Pour se faire, il faudrait qu'à chaque étape, ils conservent une trace écrite de la question posée.

▷ **Avis**

La fiche méthodologique standard pour un type d'exercice donné, en géométrie ou en algèbre, n'existe pas, tout simplement parce qu'elle mérite d'être souvent réadaptée en fonction de l'énoncé, et en fonction de la tâche demandée.

Cette réadaptation doit être effectuée par l'élève. La fiche, initialement distribuée, lui sert alors de support ou de base de réflexion. De plus, s'il parvient à accommoder cette fiche ou à s'en détacher, cela sous-entendra qu'il commence à développer sa propre démarche de recherche. Par conséquent, l'élève serait en train de devenir un élève autonome.

1.6 Conclusion

La double mission de l'école, qui est aussi celle du collège, est d'accueillir la quasi-totalité d'une classe d'âge dans un même établissement, tout en essayant de s'adapter à chacun. Face à cette mission, l'enseignant est quelques fois amené à reconsidérer les modalités pédagogiques et méthodologiques, afin de mettre l'élève au cœur de l'apprentissage et l'amener à devenir de plus en plus acteur et autonome.

Bien que l'aide de l'enseignant soit appréciable durant les heures de cours, l'apprentissage de l'autonomie doit être encouragée. Elle permet à l'élève d'être acteur, de voir l'utilité de ce qui est à faire et de mener à bien la tâche qui lui est demandée.

En proposant à mes élèves de sixième des outils, tels que le brouillon et la fiche méthodologique, tout en insistant sur la recherche des données et la reformulation, mon intention première était de les amener à réagir efficacement face à un exercice et à leur proposer des méthodes pour mieux approcher la solution. Cette démarche leur permettait aussi de prendre conscience que pour réagir efficacement, il fallait au moins essayer et non partir vaincu d'avance. Et cela, en s'efforçant de faire des dessins, de rechercher des liens avec le cours, de repérer les données essentielles dans un énoncé mathématique...

Le processus de compréhension de l'énoncé est un travail de longue haleine, qui ne peut être acquis en une année scolaire. Tout au long du cursus scolaire de l'élève, qu'il conviendra à chaque enseignant de poursuivre cette tâche, et permettre l'apprentissage de l'autonomie, à travers les différents dispositifs d'aide personnalisée, tels que les cours de soutien, les modules, les études dirigées, les cours de méthodologie…

Travail fait dans les classes antérieures sur les énoncés et consignes

A l'école primaire, en ce qui concerne les compétences transversales, nous retrouvons dans la partie « Traitement de l'information » les informations suivantes :

- Au cycle 1,l'enfant « comprend et exécute une consigne »
- Au cycle 2 il sait « utiliser un mode d'emploi, une notice ».
- Au cycle 3, il sait « sélectionner les informations utiles et les organiser logiquement ».

Dans la partie « compétences dans le domaine de la langue », sous -partie « lecture », nous retrouvons la problématique des consignes, qui se décline de la façon suivante :

- En cycle 3, l'élève doit pouvoir « exécuter une consigne »
- Dès le cycle 2, en expression écrite, l'élève est invité à rédiger des textes « prescriptifs » (règles de jeux, règles de vie, modes d'emploi…) après avoir, dès le cycle 1, produit des textes de ce type (par exemple des recettes) en les dictant au maître.

Dans l'ouvrage « Maîtrise de la langue au collège », on précise qu'en cycle 3, chaque enfant doit accéder à une attitude de lecteur autonome sur des textes injonctifs (il est sous-entendu les textes qui induisent des actions intellectuelles ou matérielles, en particulier les consignes scolaires).

Annexe 7

Bibliographie

[1] M. Baudry, D. Bessonnat, M. Laparra, F. Tourigny, La maîtrise de la langue au collège, CNDP, 1998.

[2] M.-F. Chesnais, Vers l'autonomie, Hachette Education, 1998.

[3] J.-P. Charton, C. Flinois, D. Levêque, P. Roussel, Mathématiques en sixième, Coll. Méthodes en pratique, CRDP du Nord-Pas-de-Calais, 1997.

[4] G. Chapiron, M. Mante, R. Mulet-Marquis, C. Perotin, Manuel de sixième, Collection Triangle, Edition Hatier, 2000.

[5] A. Descaves, Comprendre des énoncés, résoudre des problèmes, Hachette Education, 1992.

[6] B. Lahire, La construction de l'autonomie à l'école primaire : entre savoirs et pouvoir, 2001.

[7] A. Lanoelle, F. Perrinaud, J.-C. Rivoallan, Manuel de sixième de la collection Dimathème, Edition 2000.

[8] P. Legrand, Les maths en collège et lycée, Coll. Profession enseignant, Hachette Education, 2002.

[9] R. Bruno & L. Grosjean, Apprendre ensemble, pour une pédagogie de l'autonomie, CRDP Académie de Grenoble, 1999.

[10] J. Ravestein, Autonomie de l'élève et régulation du système didactique, 1999.

[11] J.-M. Zakhartchouk, Comprendre les énoncés et les consignes, Cahiers pédagogiques, 1999.

Chapitre 2

Histoires de groupes

Histoires de groupes

(Dominique Hoareau[1])

Résumé : Voici un document de synthèse sur les groupes finis que l'on aura plaisir à parcourir. Il regroupe du matériel utile si l'on désire passer un concours comme le CAPES ou l'agrégation, et permet de mettre en oeuvre des techniques spécifiques dans les démonstrations comme les passages aux quotients, l'utilisation du premier théorème d'isomorphisme ou l'emploi du centre d'un groupe.
Une attention particulière a été portée aux groupes de permutations. La section 2.7 propose de tirer avec remise deux éléments d'un groupe fini non commutatif G, et de calculer la probabilité $P(G)$ pour que ces deux éléments commutent entre eux (théorème de Dixon). Enfin la dernière section explicite la structure de certains groupes (groupes d'exposant 2, groupes diédraux, groupes d'ordre pair ou d'ordre $2p$ avec p premier impair, théorème de Cauchy) d'une façon toujours attayante et en établissant de nombreux parallèles.
Pas moins de 25 exercices sont proposés, le plus souvent avec une solution développée.

[1]IUT de Montpellier, domeh@wanadoo.fr.

2.1 Division euclidienne dans $\mathbb{Z}$

2.1.1 Sous-groupes de $\mathbb{Z}$, congruence dans $\mathbb{Z}$

Propriété 2.1 *Les sous-groupes de $(\mathbb{Z},+)$ sont monogènes. Plus précisément, les sous-groupes additifs de $\mathbb{Z}$ sont de la forme $n\mathbb{Z}$, où $n \in \mathbb{N}$.*

Preuve : Les $n\mathbb{Z}$ sont des sous-groupes de $\mathbb{Z}$. Réciproquement, soit H un sous-groupe de $\mathbb{Z}$, $H \neq \{0\}$. La partie $H \cap \mathbb{N}^*$ non vide de $\mathbb{N}$ a un plus petit élément qu'on appelle n. Clairement, $n\mathbb{Z} \subset H$. A présent, si $x \in H$, par division euclidienne, x s'écrit $x = nq + r$, avec $0 \leqslant r < n$. Ainsi $r = x - nq$ est dans H et dans $\mathbb{N}^*$, et $r = 0$ par statut de n. Finalement $x = nq \in n\mathbb{Z}$. D'où le résultat. On retiendra qu'un générateur d'un sous-groupe $H \neq \{0\}$ de $\mathbb{Z}$ est le plus petit entier naturel non nul qui se trouve dans H. ■

Pour a et b entiers relatifs, les sous-groupes $a\mathbb{Z} + b\mathbb{Z}$ et $a\mathbb{Z} \cap b\mathbb{Z}$ s'écrivent (de façon unique) $d\mathbb{Z}$ et $m\mathbb{Z}$ ($d, m \in \mathbb{N}$). L'entier d est appelé pgcd de a et b (notation : $d = a \wedge b$), m ppcm de a et b ($m = a \vee b$). Lorsque $d = 1$, on dit que a et b sont *premiers entre eux.*

Soit $n \in \mathbb{N}$. On définit la relation de congruence modulo n en posant :

$$x \equiv y(n) \;\Leftrightarrow\; x - y \in n\mathbb{Z}.$$

L'ensemble des classes d'équivalence est notée $\mathbb{Z}/n\mathbb{Z}$ et la classe de x modulo $n\mathbb{Z}$ est $\bar{x} = x + n\mathbb{Z}$.

Lorsque $x \equiv 0(n)$, on dit que n divise x et on note : $n \mid x$.

Propriété 2.2 *(lemme de Gauss)*
Si $a \mid bc$ et $a \wedge b = 1$, alors $a \mid c$.

Preuve : Puisque $a \wedge b = 1$, il existe u et v dans $\mathbb{Z}$ tels que $au + bv = 1$. On a alors $c(au + bv) = c$, $acu + bcv = c$. Or bc s'écrit $bc = a\gamma$ avec $\gamma \in \mathbb{Z}$. Il vient $a(cu + \gamma v) = c$ donc $a \mid c$. ■

Exercice 2.1
1) Pour $d = a \wedge b$, montrer que $d \mid a$ et $d \mid b$, et que $[d' \mid a \;;\; d' \mid b \Rightarrow d' \mid d]$.
2) Pour $m = a \vee b$, montrer que $a \mid m$ et $b \mid m$, et que $[a \mid m' \;;\; b \mid m' \Rightarrow m \mid m']$.
3) Pour $m = a \vee b$, montrer qu'il existe a' et b' dans $\mathbb{Z}$, tels que $a' \mid a$, $b' \mid b$, $a' \wedge b' = 1$ et $m = a'b'$. (On pourra décomposer a et b en produits de facteurs premiers :

$$a = \underbrace{p_1^{\alpha_1} \dots p_k^{\alpha_k}}_{a'} p_{k+1}^{\alpha_{k+1}} \dots p_j^{\alpha_j} \quad ; \quad b = p_1^{\beta_1} \dots p_k^{\beta_k} \underbrace{p_{k+1}^{\beta_{k+1}} \dots p_j^{\beta_j}}_{b'},$$

où $\alpha_i > \beta_i \geqslant 0$ si $1 \leqslant i \leqslant k$ et $0 \leqslant \alpha_i \leqslant \beta_i$ si $k+1 \leqslant i \leqslant j$.)

Un entier p est dit *premier* si ses seuls diviseurs sont p, $-p$, 1 et -1. S'il ne divise pas un entier k, alors $k \wedge p = 1$.

Propriété 2.3 $\mathbb{Z}/n\mathbb{Z} = \{\bar{0}, \bar{1}, ..., \overline{n-1}\}$.

Preuve : Pour α et α' distincts entre 0 et $n-1$, on a $\alpha' - \alpha \notin n\mathbb{Z}$, donc $\bar{\alpha} \neq \overline{\alpha}'$. Par ailleurs, pour $m \in \mathbb{Z}$, on écrit $m = nq + r$ où $q \in \mathbb{Z}$ et $0 \leqslant r \leqslant n-1$. Ainsi : $m - r \in n\mathbb{Z}$ et $\bar{m} = \bar{r}$. ■

2.1.2 Le modèle $\mathbb{Z}$

Propriété 2.4 *Tout groupe $G =< a >$ monogène et infini est isomorphe à $\mathbb{Z}$.*

Preuve : On considère le morphisme de groupes surjectif $f : n \mapsto a^n$ de $\mathbb{Z}$ sur G. Si f n'est pas injective, $\exists m, n \in \mathbb{Z} \quad m \neq n,\ a^m = a^l$, donc $a^{m-l} = 1$. On considère alors n, le plus petit entier naturel tel que $a^n = 1$. On a : $\{1, a, ..., a^{n-1}\} \subset G$ et on montre par division euclidienne l'autre inclusion $\{1, a, ..., a^{n-1}\} \supset G$. Absurde puisque G est supposé infini. ■

2.1.3 Ordre d'un élément

► Soit G un groupe, $a \in G$ et $k \in \mathbb{Z}$. Si $a^k = 1$, on dit que k est un *exposant de a*.

► On montre que l'ensemble $\mathcal{E}_a$ des exposants de a est un sous-groupe de $\mathbb{Z}$, donc $\mathcal{E}_a$ est un $\alpha\mathbb{Z}$ avec $\alpha \in \mathbb{N}$. Lorsque $\alpha \neq 0$, on dit que α est l'*ordre* de a ou que a est d'ordre α. Sinon, on dit que a est d'ordre infini.

► On envisage alors le morphisme surjectif $f : k \mapsto a^k$ de $\mathbb{Z}$ sur $< a >$ et la relation d'équivalence sur $\mathbb{Z}$ définie par :

$$\begin{aligned} k\mathcal{R}l &\Leftrightarrow f(k) = f(l) \\ &\Leftrightarrow a^{k-l} = 1 \\ &\Leftrightarrow k - l \text{ est un exposant de } a \\ &\Leftrightarrow k - l \in \mathcal{E}_a = \alpha\mathbb{Z}. \end{aligned}$$

On vérifie que $\bar{f} : \bar{k} \mapsto f(k)$ est correctement définie de $\mathbb{Z}/\alpha\mathbb{Z}$ dans $< a >$, surjective (comme f) et injective ("par construction"). Ainsi $< a >$ est infini comme $\mathbb{Z}/0\mathbb{Z} = \mathbb{Z}$ si a est d'ordre infini ($< a > \simeq \mathbb{Z}/0\mathbb{Z}$), et $\sharp(< a >) = \alpha$ si a est d'ordre $\alpha \neq 0$. On a montré :

Propriété 2.5 *Dans un groupe G, l'ordre d'un élément est égal à l'ordre du sous-groupe qu'il engendre.*

▶ On suppose que G est un groupe fini. L'ensemble $\mathcal{M}$ des entiers k tels que $(\forall a \in G \quad a^k = 1)$ est aussi un sous-groupe de $\mathbb{Z}$: en fait, $\mathcal{M} = \bigcap_{a\in G} \mathcal{E}_a$, donc $\mathcal{M}$ est $m\mathbb{Z}$ où m est le *ppcm* des ordres des éléments de G et aussi le plus petit entier naturel tel que $(\forall a \in G \quad a^m = 1)$. L'entier m est appelé *exposant de* G. Par exemple dans $\mathcal{S}_3$ (d'ordre 6), les permutations autres que 1 sont d'ordre 2 (transpositions) et d'ordre 3 (3-cycles), donc l'exposant de $\mathcal{S}_3$ est 6.

2.1.4 Groupe cyclique

Propriété 2.6 *Soit* $G =< a >= \{1, a, ..., a^{n-1}\}$ *un groupe cyclique d'ordre* n. *Alors*

1. Tout sous-groupe $H \neq \{1\}$ *de* G *est cyclique, engendré par la plus petite puissance de* a *qui se trouve dans* H.

2. Pour $0 \leqslant k \leqslant n-1$, *l'élément* a^k *est d'ordre* $\alpha = \frac{n}{k\wedge n}$. *Aussi,*

$$< a^k >= G \quad \Leftrightarrow \quad k \wedge n = 1.$$

3. Pour tout diviseur d *de* n, *il existe un et un seul sous-groupe de* G, *d'ordre* d.

Preuve : 1. On procède comme pour la caractérisation des sous-groupes de $\mathbb{Z}$. On écrit $H = \{1, a^{k_1}, ..., a^{k_j}\}$ où $0 < k_1 < ... < k_j < n$, et par division euclidienne de k_i par k_1, on montre que tout a^{k_i} est une puissance de a^{k_1}. On retiendra qu'un générateur de H est la plus petite puissance de a qui se trouve dans H.

2. On écrit $< a^k >= \{1, a^{k_1}, ..., a^{k_{\alpha-1}}\}$ (il y a α éléments dans $< a^k >$) où $0 < k_1 < ... < k_{\alpha-1} < n$. On sait que a^k et a^{k_1} engendrent le même sous-groupe. Or, par statut de α (plus petit entier naturel tel que $\left(a^{k_1}\right)^\alpha = 1$), 1 et les puissances $a^{k_1}, a^{2k_1}, ..., a^{(\alpha-1)k_1}$ de a^{k_1} (au nombre de α) sont distincts. D'où nécessairement :

$$< a^k >=< a^{k_1} >= \{1, a^{k_1}, a^{k_2} = a^{2k_1}, ..., a^{k_{\alpha-1}} = a^{(\alpha-1)k_1}\}.$$

L'entier $k_1 \times \alpha$ est un exposant de a, donc $n \mid k_1\alpha$. Or, $k_{\alpha-1} = (\alpha-1)k_1 < n$, donc $\alpha k_1 - k_1 < n$, $\alpha k_1 < n + k_1 < 2n$, d'où : $n = k_1 \times \alpha$.
Par ailleurs, $a^k \in< a^{k_1} >$, donc k s'écrit $k = k_1 \times \beta$.
Enfin, puisque $a^{k_1} \in< a^k >$, $a^{k_1} = a^{k\,s}$, $k_1 - ks$ est un exposant de a, donc s'écrit $k_1 - ks = nt$. Il vient : $k_1 - k_1\beta\ s = k_1\alpha t$, $\alpha t + \beta s = 1$, donc $\alpha \wedge \beta = 1$. Puisque par ailleurs $n = k_1\alpha$ et $k = k_1\beta$, on a $k_1 = k \wedge n$ et

$$\alpha = \frac{n}{k \wedge n}.$$

3. *Existence* : Soit d un diviseur de n. On pose $k = \frac{n}{d}$ et $x = a^k$. L'ordre de x est d'après ce qui précède : $\frac{n}{k \wedge n} = \frac{n}{k} = d$.

Unicité : Soit H un sous-groupe d'ordre d de G (cyclique), soit alors $y = a^m$ dans G tel que $H =< y >$. Pour $g = a^\alpha \in< y >$, $g^d = a^{\alpha d} = 1$, donc il existe $l \in \mathbb{Z}$ tel que $\alpha d = l\,n$, $\alpha = \frac{l\,n}{d}$. Ainsi : $g = a^\alpha = a^{\frac{l\,n}{d}} = (a^{\frac{n}{d}})^l = (a^k)^l = x^l$ donc $g \in< x >$ et $< y >\subset< x >$. Puisque $\sharp\,(< y >) = \sharp\,(< x >)$, $< x >=< y >$. ∎

Question 2.1 *Un groupe dont tous les sous-groupes sont cycliques, est-il nécessairement cyclique ?*

Exercice 2.2 *Si G et $\{1\}$ sont les seuls sous-groupes d'un groupe G non trivial, alors G est monogène, fini et d'ordre premier.*

Soit $x \in G$, $x \neq 1$. On a $< x >\neq \{1\}$, donc $< x >= G$. Par ailleurs on a $< x^2 >= \{1\}$ ou $< x^2 >=< x >$, donc $x^2 = 1$ ou bien $\exists m \in \mathbb{Z} \quad (x^2)^m = x$, i.e. $x^2 = 1$ ou $x^{2m-1} = 1$. Ainsi, G est cyclique de cardinal noté n. Soit enfin $d \mid n$ avec $1 < d < n$. Le sous-groupe $< x^{\frac{n}{d}} >$ est d'ordre $d \notin \{1, n\}$, donc $< x^{\frac{n}{d}} >\neq \{1\}$ et $< x^{\frac{n}{d}} >\neq G$. Impossible !

2.2 Théorème de Lagrange

2.2.1 Dans un groupe abélien fini

Soit G un groupe commutatif fini d'ordre n, et $g \in G$. On pose :

$$a = \prod_{x \in G} x.$$

La translation $t_g : x \mapsto gx$ est bijective et puisque G est commutatif, le produit dans G ne dépend pas de l'ordre des facteurs. Ainsi :

$$a = \prod_{x \in G} t_g(x) = \prod_{x \in G} gx$$

et toujours par commutativité de G, $a = g^n a$. En définitive :

$$\forall g \in G \qquad g^n = 1.$$

2.2.2 Relation modulo un sous-groupe

Soit G un groupe. Si H est un sous-groupe de G, on définit sur G la relation de *congruence (à gauche) modulo H* en posant : $g\mathcal{R}_H g' \Leftrightarrow g'^{-1}g' \in H$. Pour

$g \in G$, la classe de g est $\bar{g} = \{gh \; ; \; h \in H\}$, qu'on note aussi gH. L'espace des classes est noté G/H.
On suppose que G est fini. Pour $g \in G$, on vérifie que l'application $h \mapsto gh$ de H dans gH est une bijection, ce qui assure que toutes les classes ont le même cardinal $\circ(H)$.
Si on note $[G : H]$ le nombre de classes d'équivalence modulo $\mathcal{R}_H$ (indice de H dans G), on peut écrire :

$$\circ(G) = [G : H] \times \circ(H).$$

On vient d'établir le *théorème de Lagrange* :

Propriété 2.7 *Si G est un groupe fini et si H est un sous-groupe de G, alors l'ordre de H et son indice dans G divisent l'ordre de G.*

Exercice 2.3 *Montrer que si G est un groupe de cardinal impair, alors :*

$$\forall x \in G \quad \exists ! \, y \in G \quad x = y^2.$$

Solution : Pour tout x dans G, l'ordre du sous-groupe $< x >$ engendré par x divise $\circ(G) = 2k + 1$, donc $x^{2k+1} = 1$, et

$$x = x.x^{2k+1} = (x^{k+1})^2 = y^2$$

en posant $y = x^{k+1}$. L'unicité se montre en notant que $y^{2k+1} = z^{2k+1} = 1$ pour tous $y, z \in G$, et que par conséquent si $x = y^2 = z^2$, on obtient

$$y^{2k+1} = z^{2k+1} \;\Rightarrow\; x^k.y = x^k.z \;\Rightarrow\; y = z.$$

Exercice 2.4 *On désigne par $\mathcal{S}_n$ le groupe des bijections de $\{1, ..., n\}$. On veut calculer son ordre β_n. On définit sur $\mathcal{S}_n$ une relation d'équivalence en posant : $\sigma \mathcal{R} \sigma'$ si, et seulement si, $\sigma(n) = \sigma'(n)$.*

1) Démontrer que $\mathcal{R}$ est la relation d'équivalence modulo le sous-groupe $H_n = \{\sigma \in \mathcal{S}_n \; ; \; \sigma(n) = n\}$ de $\mathcal{S}_n$.

2) Montrer que H_n est en bijection avec $\mathcal{S}_{n-1}$ et que $\mathcal{S}_n/H_n$ est en bijection avec $\{1, ..., n\}$.

3) En déduire par récurrence que $\beta_n = \circ(\mathcal{S}_n) = n!$.

Question 2.2 *Si G est un groupe fini d'ordre n et si m est un diviseur de n, G possède-t-il un sous-groupe d'ordre m ? La réponse est positive si G est cyclique.*

Réponse : Pas forcément, comme on le voit dans la remarque qui suit la Propriété 2.20.

Corollaire 2.1 *L'ordre d'un groupe fini est un exposant de chacun de ses éléments : si G est un groupe fini d'ordre n, alors $g^n = 1$ pour tout $g \in G$.*

Preuve : Soit $g \in G$. L'ordre de g est l'ordre du sous-groupe $< g >$, donc divise l'ordre de G, d'où le résultat. ■

Corollaire 2.2 *Un groupe G d'ordre premier p est cyclique.*

Preuve : Soit $x \in G$, $x \neq 1$. On a $\circ(x) > 1$ et $\circ(x)$ est un diviseur de p, donc avec p premier, $\circ(x) = p$. Ainsi $G =< x >= \{1, x, .., x^{p-1}\}$. ■

Exercice 2.5 *Trouver les sous-groupes de $\mathcal{S}_3$.*

Solution : On étiquette les éléments de $\mathcal{S}_3$. On trouve l'identité Id, les transpositions (permutations d'ordre 2) $\tau_1 = (2, 3)$, $\tau_2 = (1, 3)$, $\tau_3 = (1, 2)$, et les 3-cycles $r = (1, 2, 3)$, $\rho = r^{-1} = (1, 3, 2)$. Si H est un sous-groupe propre de $\mathcal{S}_3$, son ordre est 2 ou 3, donc H est cyclique. Ainsi $H =< \tau_1 >$, ou $< \tau_2 >$, ou $< \tau_3 >$, ou $< r >=< \rho >$.
$\mathcal{S}_3$ nous fournit un exemple de groupe non cyclique (et même non commutatif) dont tous les sous-groupes propres sont cycliques.

Corollaire 2.3 *(Formule des indices) Soient S et T deux sous-groupes de G tels que : $S \subset T$. On a : $[G : S] = [G : T]\ [T : S]$.*

Preuve : • *Première méthode* : on écrit

$$[G : S] = \frac{\circ(G)}{\circ(S)} = \frac{[G : T]\ \circ(T)}{\circ(S)} = [G : T]\ [T : S].$$

• Seconde méthode (*preuve directe*) :

- Si $xS = x'S$, alors $x^{-1}x' \in S$, donc $x^{-1}x' \in T$ car $S \subset T$, ce qui assure que $xT = x'T$. On définit ainsi correctement une application ϕ de G/S vers G/T en posant $\phi(xS) = xT$.

- Soit $\mathcal{R}$ la relation d'équivalence associée à ϕ, définie sur G/S par :

$$xS\,\mathcal{R}\,x'S \quad \Leftrightarrow \quad \phi(xS) = \phi(x'S).$$

On a : $xS\,\mathcal{R}\ x'S \ \Leftrightarrow\ x\mathcal{R}_T x'$, donc le nombre de classes d'équivalence pour $\mathcal{R}$ est égal à $[G : T]$.

- On a : $\overline{xS} = \{x'S\ ;\ x' \in xT\} = \{xtS\ ;\ t \in H\}$. On vérifie que l'application $tS \mapsto xtS$ de T/S dans $\overline{xS}$ est bijective, ce qui assure que chaque classe $\overline{xS}$ possède $[T : S]$ éléments. On conclut alors aisément. ■

2.2.3 Congruence dans $\mathbb{Z}$

▶ Pour $n \in \mathbb{N}$, les classes d'équivalence de $\mathbb{Z}/n\mathbb{Z}$ sont notées $\bar{k}$ (où $k \in \mathbb{Z}$) et l'on a $\mathbb{Z}/n\mathbb{Z} = \{\bar{0}, ..., \overline{n-1}\}$. Comme $(\mathbb{Z}, +)$ est commutatif, on montre facilement que :

$$\bar{k} = \bar{l} \quad et \quad \bar{l} = \bar{l'} \quad \Rightarrow \quad \overline{k+l} = \overline{k'+l'}.$$

On dit que la relation de congruence est compatible avec la loi additive de $\mathbb{Z}$. La loi de composition interne $(\bar{k}, \bar{l}) \mapsto \overline{k+l}$ est ainsi correctement définie, et donne à $\mathbb{Z}/n\mathbb{Z}$ une structure de groupe commutatif.

▶ Si $G =< a >$ désigne un groupe monogène, on envisage le morphisme surjectif $f_a : n \mapsto a^n$ de $\mathbb{Z}$ sur G et la relation d'équivalence sur $\mathbb{Z}$:

$$\begin{array}{rcl} x\mathcal{R}_a y & \Leftrightarrow & f_a(x) = f_a(y) \\ & \Leftrightarrow & f_a(x)f_a(-y) = 1 \\ & \Leftrightarrow & f_a(x-y) = 1 \\ & \Leftrightarrow & x - y \in ker(f_a). \end{array}$$

Or $ker(f_a)$ est un sous-groupe de $\mathbb{Z}$, donc l'ensemble des classes d'équivalence pour $\mathcal{R}_a$ est un $\mathbb{Z}/n\mathbb{Z}$. Deux cas se présentent :

a) $n = 0$, f_a est alors injective et G est isomorphe à $\mathbb{Z}$.

b) $n \neq 0$, on vérifie facilement que $\widetilde{f_a} : \bar{x} \mapsto f_a(x)$ est correctement définie de $\mathbb{Z}/n\mathbb{Z}$ dans G, surjective (comme f_a), injective et réalise un morphisme de $\mathbb{Z}/n\mathbb{Z}$ sur G.

On a montré :

Propriété 2.8 *Soit G un groupe monogène.*
1) Si G est infini, G est isomorphe à $\mathbb{Z}$.
2) Si G est fini d'ordre n, alors G est isomorphe à $\mathbb{Z}/n\mathbb{Z}$.

Remarque : Si G est cyclique, on peut penser qu'il y a autant d'isomorphismes de $\mathbb{Z}/n\mathbb{Z}$ sur G, que de générateurs de G.

Exercice 2.6 *Si un groupe G possède exactement 3 sous-groupes, alors G est cyclique d'ordre p^2, avec p premier.*

Solution : Soit $x \in G$, $x \neq 1$. Le sous-groupe $< x >$ est différent de $\{1\}$ donc deux cas se présentent :

a) Si $< x >= G$, alors $< x >$ possède un seul sous-groupe propre non trivial. Puisque $\mathbb{Z}$ possède une infinité de sous-groupes propres, $< x >$ n'est pas

isomorphe à $\mathbb{Z}$, est donc cyclique. Son ordre possède alors un unique diviseur propre et s'écrit nécessairement p^2, avec p premier.

b) Si $< x > \neq G$, $\{1\}$, $< x >$ et G sont les trois sous-groupes de G. Soit $y \in G\setminus < x >$. On a $< y > \neq \{1\}$ et $< x > \neq < y >$, donc $< y >= G$. On est alors ramené au cas précédent.

Question 2.3 *Il n'y a qu'un seul modèle de groupe d'ordre n premier : $\mathbb{Z}/n\mathbb{Z}$. Mais peut-on affirmer que n est premier lorsqu'il n'y a qu'un modèle de groupe d'ordre n ?*

► On vérifie que la congruence modulo n est également compatible avec la loi multiplicative de $\mathbb{Z}$, ce qui permet de définir une loi multiplicative sur $\mathbb{Z}/n\mathbb{Z}$. Muni de ces deux lois, $\mathbb{Z}/n\mathbb{Z}$ possède une structure d'anneau.

► On rappelle que les inversibles de $\mathbb{Z}/n\mathbb{Z}$ (éléments ayant un symétrique pour la loi multiplicative) forment un groupe noté $\sqcup(\mathbb{Z}/n\mathbb{Z})$ ou $(\mathbb{Z}/n\mathbb{Z})^*$, et que

$$\bar{k} \in (\mathbb{Z}/n\mathbb{Z})^* \quad \Leftrightarrow \quad k \wedge n = 1.$$

Propriété 2.9 *(Lemme d'Euclide)*
Si p est premier et si $p \mid ab$, alors $p \mid a$ ou $p \mid b$.

2.2.4 Indicatrice d'Euler

Pour $n \geqslant 1$, on note $\varphi(n)$ le nombre des entiers $k \in \{1, ..., n\}$ premiers avec n. C'est aussi l'ordre du groupe multiplicatif $\sqcup(\mathbb{Z}/n\mathbb{Z})$, ou encore le nombre des $\bar{k}$ du groupe $(\mathbb{Z}/n\mathbb{Z}, +)$ tels que le sous-groupe $< \bar{k} >= \{j\bar{k}\ ;\ j \in \mathbb{Z}\}$ engendré par $\bar{k}$ soit égal à $\mathbb{Z}/n\mathbb{Z}$; C'est aussi le nombre des générateurs d'un groupe cyclique d'ordre n, ou enfin le nombre des éléments de $(\mathbb{Z}/n\mathbb{Z}, +)$ qui sont d'ordre exactement n.
Lorsque $n = p$ est premier, $\mathbb{Z}/p\mathbb{Z}$ est un corps et $\varphi(p) = p - 1$. Le théorème de Lagrange donne :

Propriété 2.10 *(Petit théorème de Fermat)*
Pour p premier et $k \neq 0\ (p)$, on a : $k^{p-1} = 1$.

Application 2.1 *Soit p premier. Dans l'anneau $(\mathbb{Z}/p\mathbb{Z})[X]$, on a l'identité :*

$$X^p - X = \prod_{k=0}^{p-1} (X - k).$$

Conséquences :
(1) $(p-1)! = -1\ (p)$.

(2) $(X+1)^p = X^p + 1$ *dans l'anneau* $(\mathbb{Z}/p\mathbb{Z})[X]$.

(3) *Pour* $0 < k < p$, $C_p^k = 0$ (p).

On constate en effet que le polynôme $X^p - X$ *unitaire de degré* p *admet les* p *éléments* k *du corps* $\mathbb{Z}/p\mathbb{Z}$ *comme racines, donc est le produit des* $X - k$. *Pour* (1), *on identifie le coefficient de* X *dans l'identité précédente. Pour* (2), *on écrit*

$$\begin{aligned} X^p - X &= \prod_{k \in \mathbb{Z}/p\mathbb{Z}} (X - k) \\ &= \prod_{k \in \mathbb{Z}/p\mathbb{Z}} (X - (k-1)) \\ &= \prod_{k \in \mathbb{Z}/p\mathbb{Z}} ((X+1) - k) = (X+1)^p - (X+1). \end{aligned}$$

Enfin, la nullité des coefficients binomiaux C_p^k $(0 < k < p)$ *s'obtient en développant* (2).

▶ Comment calculer $\varphi(n)$, pour tout $n \in \mathbb{N}^*$?

1) Pour p premier, $\varphi(p) = p - 1$.

2) Pour p premier et $\alpha \in \mathbb{N}$, $\varphi(p^\alpha) = p^\alpha - p^{\alpha-1}$.

3) Si m et n sont deux entiers premiers entre eux, on démontre que $\mathbb{Z}/mn\mathbb{Z}$ et $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}$ sont des groupes (et même des anneaux) isomorphes (Lemme 2.1). Conséquence : $\varphi(mn) = \varphi(m)\ \varphi(n)$ dès que $m \wedge n = 1$. On dit que la fonction indicatrice d'Euler φ est *multiplicative* au sens de l'arithmétique.

4) Si n se décompose en facteurs premiers sous la forme $n = p_1^{\alpha_1} ... p_i^{\alpha_i}$, alors

$$\varphi(n) = n \prod_{k=1}^{i} (1 - \frac{1}{p_k}).$$

Lemme 2.1 *$\mathbb{Z}/mn\mathbb{Z}$ et $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}$ sont des anneaux isomorphes.*

Preuve : Puisque m et n sont premiers entre eux, on choisit $m', n' \in \mathbb{Z}$ tels que $mm' + nn' = 1$. Pour p, p', q, q' appartenant à $\mathbb{Z}$, $p = p'(n)$ et $q = q'(m)$ impliquent $p - p' \in n\mathbb{Z}$, $q - q' \in m\mathbb{Z}$, ce qui donne : $(p-p')mm' + (q-q')nn' \in mn\mathbb{Z}$ ou $pmm' + qnn' = p'mm' + q'nn'(mn)$. On définit donc correctement une application ϕ de $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}$ dans $\mathbb{Z}/mn\mathbb{Z}$ en posant : $\phi(\bar{p}, \bar{q}) = \overline{pmm' + qnn'}$. On vérifie facilement que $\phi((\bar{p}, \bar{q}) + (\bar{p}', \bar{q}')) = \phi(\bar{p}, \bar{q}) + \phi(\bar{p}', \bar{q}')$, ce qui fait de ϕ un morphisme de groupe. A présent, si $\phi(\bar{p}, \bar{q}) = 0$, alors $pmm' + qnn' \in mn\mathbb{Z}$,

donc $pmm' \in n\mathbb{Z}$ et comme $m \wedge n = m' \wedge n = 1$, on a $p \in n\mathbb{Z}$. De même, on vérifie que $q \in m\mathbb{Z}$, donc ϕ est injective. Puisque $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}$ et $\mathbb{Z}/nm\mathbb{Z}$ ont le même cardinal, finalement, on obtient bien : $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z} \simeq \mathbb{Z}/nm\mathbb{Z}$. ■

▶ Voici une égalité classique :

Propriété 2.11 $n = \sum_{d|n} \varphi(d)$.

Preuve : • *Première méthode.* On utilise le résultat suivant (une réciproque partielle à Lagrange) :

> *Lemme* : Si $G =< a >$ est un groupe cyclique d'ordre n, pour tout diviseur d de n, il existe un et un seul sous-groupe de G, d'ordre d.

On définit sur G une relation d'équivalence de la façon suivante : pour x et y dans G, $x\mathcal{R}y$ si, et seulement si, il existe d diviseur de n tel que $o(x) = o(y) = d$. Avec la partie *existence* du lemme, il y a autant de classes d'équivalence que de diviseurs de n. Avec la partie *unicité*,

$$x\mathcal{R}y \;\Leftrightarrow < x >=< y >,$$

donc la classe de x est l'ensemble des générateurs de $< x >$, au nombre de $\varphi(d)$. Ainsi $n = \sum_{d|n} \varphi(d)$.

• *Deuxième méthode.* On raisonne par récurrence sur le nombre l d'entiers premiers de la décomposition de n. Si $l = 1$, n s'écrit $n = p^q$ avec p premier et $q \in \mathbb{N}^*$. Les diviseurs de n sont 1, p, ..., p^{q-1}, p^q, et

$$\begin{aligned}\sum_{d|n} \varphi(d) &= 1 + (p-1) + (p^2 - p) + ... + (p^{q-1} - p^{q-2}) + (p^q - p^{q-1}) \\ &= p^q \;=\; n.\end{aligned}$$

On suppose l'égalité vraie au rang l, et on envisage un entier

$$n = \prod_{1 \le i \le l+1} p_i^{\alpha_i} = \prod_{1 \le i \le l} p_i^{\alpha_i} \times p_{l+1}^{\alpha_{l+1}}.$$

Les diviseurs de n sont les diviseurs de $\prod_{1 \le i \le l} p_i^{\alpha_i}$ dont l'ensemble est noté $\mathcal{D}_l$ et les $d_l \times p_{l+1}^{\alpha}$ où $d_l \in \mathcal{D}_l$, $1 \le \alpha \le \alpha_{l+1}$, et $d_l \wedge p_{l+1}^{\alpha} = 1$. Par multiplicativité de la fonction indicatrice d'Euler et avec quelques factorisations judicieuses, on peut écrire :

$$\sum_{d|n} \varphi(d) = \sum_{d \in \mathcal{D}_l} \varphi(d) + \sum_{k=1}^{\alpha_{l+1}} \left(\varphi(p_{l+1}^k) \sum_{d \in \mathcal{D}_l} \varphi(d) \right),$$

puis

$$\begin{aligned}\sum_{d|n}\varphi(d) &= \sum_{d\in\mathcal{D}_l}\varphi(d)\times\left(1+\sum_{k=1}^{\alpha_{l+1}}\varphi(p_{l+1}^k)\right)\\ &= \left(\prod_{1\le i\le l}p_i^{\alpha_i}\right)\left(1+(p_{l+1}^{\alpha_{l+1}}-1)\right) = n. \blacksquare\end{aligned}$$

2.2.5 Groupe $(\mathbb{Z}/p\mathbb{Z})^*$ quand p est premier

Nous aurons besoin des résultats suivants :

Lemme 2.1 *Soit G un groupe fini, a et b deux éléments de G d'ordre α et β. On suppose que $ab = ba$ et $\alpha \wedge \beta = 1$. Alors $\circ(ab) = \alpha\beta$.*

Preuve : Donnons seulement les grandes étapes de la preuve. On note γ l'ordre de ab. Alors :

- $< a > \cap < b >= \{1\}$. (On pourra calculer x^α pour $x \in< a > \cap < b >$.)
- $\gamma \mid \alpha\beta$.
- $\alpha \mid \gamma$ et $\beta \mid \gamma$. (On remarquera que $(ab)^\gamma = 1$.)
- $\gamma = \alpha\beta$. ■

Lemme 2.2 *Dans un groupe G commutatif et fini, l'ensemble des ordres est stable par ppcm. Ainsi, l'exposant m de G défini comme le ppcm des ordres des éléments de G, est aussi le plus grand des ordres des éléments de G, ce qui s'écrit : $m = \max\{\circ(x)\,;\, x \in G\}$.*

Preuve : Soit x et y deux éléments de G d'ordre $\tilde\alpha$ et $\tilde\beta$. On prend α et β dans $\mathbb{Z}$ tels que $\alpha \mid \tilde\alpha$, $\beta \mid \tilde\beta$, $\alpha \wedge \beta = 1$ et $\tilde\alpha \vee \tilde\beta = \alpha\beta$. On pose $a = x^{\tilde\alpha/\alpha}$ et $b = y^{\tilde\beta/\beta}$. L'ordre de a est α, tandis que l'ordre de b est β. Le Lemme 2.1 montre que ab est d'ordre $\alpha\beta$, c'est-à-dire $\tilde\alpha \vee \tilde\beta$. ■

Pour p premier, notons indifféremment $\sqcup(\mathbb{Z}/p\mathbb{Z})$ ou $(\mathbb{Z}/p\mathbb{Z})^*$ le groupe multiplicatif de l'anneau $\mathbb{Z}/p\mathbb{Z}$. Alors :

Propriété 2.12 *Si p est premier, $(\mathbb{Z}/p\mathbb{Z})^*$ est cyclique.*

Preuve : On veut exhiber un élément de $G = (\mathbb{Z}/p\mathbb{Z})^*$ d'ordre $p-1$. Un candidat naturel est un élément $g \in G$ tel que :

$$\circ(g) = \max\{\circ(x)\,;\, x \in G\}.$$

On pose $m = \max\{\circ(x)\,;\, x \in G\}$. La stratégie est la suivante : vérifier que tout x de G a un ordre l qui divise m. Ainsi, tous les éléments de G seront racines du polynôme $X^m - 1$ de $(\mathbb{Z}/p\mathbb{Z})[X]$, ce qui impose $p-1 \leqslant m$ et puisque par ailleurs m divise $p-1$, $m = p-1$.
C'est le Lemme 2.2 qui nous permet de conclure : m est le ppcm des ordres des éléments de G, donc est un multiple de $\circ(x)$ quel que soit x. ■

2.3 Conjugaison and co

Dans un groupe commutatif, a étant fixé dans G, pour tout $g \in G$, on a $ga = ag$ donc gag^{-1} vaut a. Dans un groupe non commutatif, gag^{-1} n'est plus nécessairement égal à a, et la diversité des résultats pour gag^{-1} quand g parcourt G, est un indicateur de la non commutativité de G. Par ailleurs, gag^{-1} même distinct de a reste proche de a. Ainsi par exemple, les exposants de a sont exactement les exposants de gag^{-1}, et $\circ(a) = \circ(gag^{-1})$.

2.3.1 Sous-groupe distingué

Deux éléments a et b sont *conjugués* ou "du même type" dans G si, et seulement si, il existe $g \in G$ tel que $b = gag^{-1}$. On définit ainsi une relation d'équivalence appelée conjugaison, qui permet de classer les éléments de G par affinité.

Exemples : a) Dans $\mathbb{R}^3$ euclidien, si s est une symétrie orthogonale par rapport à un plan P et si f est une isométrie de $\mathbb{R}^3$, alors $f\,s\,f^{-1}$ est la symétrie orthogonale par rapport au plan $f(P)$.

b) Deux matrices conjuguées de $GL(n, \mathbb{R})$ représentent le même automorphisme linéaire de $\mathbb{R}^n$, mais dans des bases différentes.

c) Dans le groupe symétrique $\mathcal{S}_n$ de degré n, si σ est un p-cycle, alors σ s'écrit $\sigma = (a_1, a_2, ..., a_p)$ et pour $\tau \in \mathcal{S}_n$, $\tau\,\sigma\,\tau^{-1}$ est le p-cycle $(\tau(a_1), \tau(a_2), ..., \tau(a_p))$.

La classe d'équivalence de $a \in G$ est appelée *orbite de a* et est notée $\mathcal{O}_a$.

Exercice 2.7 *On étiquette les éléments de* $\mathcal{S}_3$ *:* Id, $\tau_1 = (2,3)$, $\tau_2 = (1,3)$, $\tau_3 = (1,2)$, $r = (1,2,3)$ *et* $\rho = r^{-1} = (1,3,2)$. *Montrer à la main que* $\mathcal{S}_3$ *possède trois classes de conjugaison :* $\{Id\}$, $\{\tau_1, \tau_2, \tau_3\}$ *et* $\{r, \rho\}$.

Exercice 2.8 *Dans* $\mathcal{S}_3$*, deux éléments de même ordre sont conjugués. Est-ce vrai dans n'importe quel groupe ?*

Solution : Dans $\mathcal{S}_4$, considérer $(1,2)$ et $(1,2)(3,4)$, d'ordre 2 et de parité différente.

Exercice 2.9 *Quels sont les groupes commutatifs G dans lesquels les éléments de même ordre sont conjugués.*

Solution : Le groupe trivial $G = \{1\}$ convient. Soit $G \neq \{1\}$ commutatif vérifiant la propriété. Puisque toutes les orbites dans G abélien sont ponctuelles, deux éléments de même ordre sont identiques. Soit $a \in G$, $a \neq 1$ et $\alpha \geqslant 2$ son ordre. Dans le sous-groupe $< a >$, le nombre d'éléments d'ordre α est $\varphi(\alpha)$. Par nécessité, $\varphi(\alpha) = 1$. Ainsi, $\alpha \in \{1, 2\}$, et puisque $a \neq 1$, son ordre est nécessairement $\alpha = 2$. On a montré que G a 2 éléments, donc $G \equiv \mathbb{Z}/2\mathbb{Z}$. Réciproquement, $\mathbb{Z}/2\mathbb{Z}$ convient.

Un sous-groupe de G est dit *distingué* (ou *normal*, ou *invariant*) dans G, et on note $H \lhd G$, s'il contient la classe de conjugaison de chacun de ses points, i.e.

$$\forall a \in H \quad \forall g \in G \quad gag^{-1} \in H.$$

G et $\{1\}$ sont toujours distingués dans G.

Exercice 2.10 *Dans le groupe des bijections de $\mathbb{R}$ sur $\mathbb{R}$, le sous-groupe des bijections monotones est-il distingué ?*

Solution : On pose $\phi(x) = 1/x$ si $x \neq 0$, $\phi(0) = 0$, ce qui fait de ϕ une bijection de $\mathbb{R}$ sur $\mathbb{R}$, et on envisage l'application affine croissante $f : x \mapsto x+1$. On a $\phi \circ f \circ \phi^{-1}(-1) = 0$, $\phi \circ f \circ \phi^{-1}(0) = 1$, $\phi \circ f \circ \phi^{-1}(1) = 1/2$. La conclusion est aisée.

Question 2.4 *Si G est commutatif, toutes les classes de conjugaison sont des singletons et tout sous-groupe est distingué dans G. Existe-t-il des groupes non commutatifs dont tous les sous-groupes sont distingués ?*

Voici un résultat "générateur" d'exemples de sous-groupes distingués (et facile à prouver) :

Propriété 2.13 *Si G et G' sont deux groupes et f un morphisme de G dans G', alors $ker(f)$ est distingué dans G.*

Exemples : a) On considère le morphisme signature ε surjectif de $\mathcal{S}_n$ sur $\{-1, 1\}$ (qui envoie toute transposition sur -1). Son noyau, appelé *groupe alterné de degré n* et noté $\mathcal{A}_n$, est distingué. La relation d'équivalence sur $\mathcal{S}_n$ définie par :

$$\sigma \mathcal{R} \tau \quad \Leftrightarrow \quad \varepsilon(\sigma) = \varepsilon(\tau)$$

est la relation de congruence modulo $ker(\varepsilon)$ et on définit à bon droit l'application $\overline{\varepsilon} : \mathcal{S}_n/\mathcal{A}_n \to \{-1, 1\}$, injective (par construction), surjective (comme ε). Par cette bijection, $\circ(\mathcal{A}_n) = n!/2$.

b) Par le morphisme surjectif déterminant de $GL(n,\mathbb{C})$ (resp $O(n,\mathbb{R})$) sur $\mathbb{C}^*$ (resp $\mathbb{R}^*$), le groupe spécial linéaire $SL(n,\mathbb{C})$ ($SO(n,\mathbb{R})$) sont distingués.

c) H étant distingué dans G, peut-on réaliser H comme noyau d'un morphisme de groupes de source G ?

Exercice 2.11 *Avec des notations évidentes, a-t-on $H \lhd K \lhd G \Rightarrow H \lhd G$?*

Solution : On considère le groupe $GA(\mathbb{R})$ des bijections affines de $\mathbb{R}$ formé des applications $f_{a,b} : x \mapsto ax+b$ avec $a \in \mathbb{R}^*$, $b \in \mathbb{R}$.
Soit $T \equiv \mathbb{R}$ le sous-groupe des translations et $T_{\mathbb{Z}}$ le sous-groupe des $x \mapsto x+n$ ($n \in \mathbb{Z}$). Puisque T est commutatif, $T_{\mathbb{Z}} \lhd T$. Par ailleurs, T est le noyau du morphisme $f_{a,b} \mapsto a$ de $GA(\mathbb{R})$ sur $(\mathbb{R}^*, \times)$, donc T est distingué dans $GA(\mathbb{R})$. Enfin, l'égalité

$$f_{a,b}f_{1,n}{f_{a,b}}^{-1} = f_{a,b}f_{1,n}f_{\frac{1}{a},\frac{-b}{a}} = f_{1,an}$$

valable pour $(a,b) \in \mathbb{R}^* \times \mathbb{R}$ et $n \in \mathbb{Z}$, montre que $T_{\mathbb{Z}} \not\lhd GA(\mathbb{R})$ puisqu'on peut choisir $a \in \mathbb{R}^*$ et $n \in \mathbb{Z}$ tels que $an \notin \mathbb{Z}$.

Exercice 2.12 *Si G est un groupe d'ordre pair $\circ(G) = 2n$, et si H est un sous-groupe de G d'ordre n, alors H est distingué dans G.*

Solution : Le nombre de classes de congruence modulo H est 2 selon le théorème de Lagrange. Soit $x \notin H$. H et xH constituent une partition de G. Pour $a \in H$,

a) Si $g \in H$, $gag^{-1} \in H$ puisque H, comme tout sous-groupe de G, est stable pour la loi de composition interne.

b) Si $xax^{-1} \notin H$, xax^{-1} s'écrit xh, donc $ax^{-1} \in H$, $x^{-1} \in H$, $x \in H$ (stabilité de H pour le passage à l'inverse). Contradiction. On vient de vérifier que $xax^{-1} \in H$ pour tout $a \in H$.

c) Si $g = xh$ avec $h \in H$, $gag^{-1} = x(hah^{-1})x^{-1}$, qui appartient bien à H d'après ce qui précède.

On peut montrer mieux :

Propriété 2.14 *Si H est un sous-groupe d'un groupe fini G, d'indice égal au plus petit diviseur premier de $\circ(G)$, alors H est distingué dans G.*

Preuve : On va réaliser H comme noyau d'un "bon" morphisme de groupes. On note Q l'ensemble quotient $G/H = \{xH; x \in G\}$. Pour g fixé dans G et pour x et y dans G,

$$x \equiv y\ (H) \quad \Leftrightarrow \quad gx \equiv gy\ (H),$$

donc on définit correctement l'application injective $\phi_g : \bar{x} \mapsto \overline{gx}$ de Q dans Q. On envisage alors le morphisme de groupes $\Phi : g \mapsto \phi_g$ de G dans le groupe $\mathcal{S}$ des permutations de Q (action de G sur Q par translation). Pour $g \in G$,

$$g \in ker(\Phi) \;\Leftrightarrow\; \Phi(g) = Id \;\Rightarrow\; gH = H \;\Leftrightarrow\; g \in H,$$

donc $ker(\Phi) \subset H$. Par ailleurs,

$$\frac{\circ(G)}{\circ(ker(\Phi))} = \underbrace{\frac{\circ(G)}{\circ(H)}}_{p} \times \underbrace{\frac{\circ(H)}{\circ(ker(\Phi))}}_{\alpha\in\mathbb{N}^*}$$

et

$$\frac{\circ(G)}{\circ(ker(\Phi))} = \circ(Im(\Phi)) = \alpha p$$

divise $\circ(\mathcal{S}) = p! = (p-1)! \times p$, donc $\alpha \mid (p-1)!$. Ainsi, si $\alpha \neq 1$, on choisit (et c'est possible) un diviseur premier de α qui est dans $\{1, ..., p-1\}$. Ce diviseur, qui est aussi un diviseur premier de $\circ(G) = \circ(ker(\Phi)) \times \alpha p$, est supérieur à p (plus petit diviseur premier de $\circ(G)$) : contradiction et $\alpha = 1$. On a donc $H = ker(\Phi)$. ■

Remarque : Peut-on omettre la précision "plus petit" ? Non ! Dans $\mathcal{S}_3$, considérer $\sigma = (1,2)$ et $\tau = (1,3)$. Le sous-groupe $<\sigma>$ est d'ordre 2 (et d'indice 3) et $\tau \circ \sigma \circ \tau^{-1} = (2,3) \notin <\sigma>$ montre que $<\sigma>$ n'est pas distingué dans $\mathcal{S}_3$.

Exercice 2.13 *Donner une condition nécessaire et suffisante pour qu'un sous-groupe $<\sigma>$ de G d'ordre 2 soit distingué dans G.*

Solution : La CNS est $\sigma \in Z(G)$.

Exercice 2.14 *L'image directe (par un morphisme de groupes) d'un sous-groupe distingué est-elle distinguée ?*

Solution : Non. Considérer $\tau \in < (1,2) > \subset \mathcal{S}_3 \mapsto \tau \in \mathcal{S}_3$.

On dit qu'un groupe G est *simple* lorsque ses seuls sous-groupes distingués sont $\{1\}$ et G. Les groupes cycliques d'ordre premier sont simples, tandis que les groupes de permutations $\mathcal{S}_n$ de degrés $n \geqslant 3$ ne sont pas simples.

2.3.2 Centre d'un groupe

▶ **Vu comme noyau** : on peut vérifier que, pour tout $a \in G$, $\sigma_a : g \mapsto aga^{-1}$ est un automorphisme de G (dit *automorphisme intérieur*), et que $Int : a \mapsto \sigma_a$ est un morphisme de G dans le groupe $Aut(G)$ des automorphismes de G. Son noyau

$$Z(G) = \{z \in G \,/\, \forall g \in G \quad gz = zg\}$$

et est appelé *centre de G*.

▶ **Centre du groupe linéaire et du groupe orthogonal** : Soit $\mathbb{K}$ un corps commutatif et E un $\mathbb{K}$-espace vectoriel. On note $\mathcal{L}(E)$ l'algèbre des endomorphismes de E et $GL(E)$ le groupe linéaire de E. Lorsque E est un espace euclidien ($\mathbb{R}$-espace vectoriel de dimension finie muni d'un produit scalaire), on note $\mathcal{O}(E)$ le groupe orthogonal de E.
On a la caractérisation suivante des homothéties :

Lemme 2.3 (1) *Un endomorphisme f de E est une homothétie si, et seulement si,* (2) *pour tout $x \in E$, la famille $(x, f(x))$ est liée.*

Preuve : L'implication (1) $\Rightarrow$ (2) est immédiate. On suppose à présent que pour tout $x \in E$, il existe un scalaire λ_x tel que $f(x) = \lambda_x x$. On choisit $x_0 \in E$, $x_0 \neq 0$, et on montre que $\lambda_x = \lambda_{x_0}$ pour tout $x \in E$. On distingue deux cas : $x \in \mathbb{K}x_0$ et (x, x_0) libre. Si $x = \mu x_0$ avec $\mu \in \mathbb{K}$, $f(x) = \mu\, \lambda_{x_0} x_0 = \lambda_{x_0}\, x$, ce qui conduit à $\lambda_x = \lambda_{x_0}$. Sinon,

$$\lambda_{x+x_0}(x + x_0) = f(x + x_0) = f(x) + f(x_0) = \lambda_x x + \lambda_{x_0} x_0$$

et par liberté de (x, x_0), $\lambda_{x+x_0} = \lambda_x = \lambda_{x_0}$. ■

Et voici deux résultats importants :

Propriété 2.15 *On suppose E de dimension finie. Le centre de $GL(E)$ est le sous-groupe des homothéties de rapport non nul.*

Preuve : Soit $f \in GL(E)$ telle que f commute avec tout élément de $GL(E)$. Si f n'est pas une homothétie, alors on peut trouver un vecteur $x \in E$ tel que $(e_1 = x, e_2 = f(x))$ soit libre. On complète (e_1, e_2) en une base $(e_1, e_2, e_3, ..., e_n)$. Soit $g \in \mathcal{L}(E)$ définie en posant $g(e_1) = g(x) = x = e_1$, $g(e_2) = x + f(x) = e_1 + e_2$, et $g(e_i) = e_i$ pour $3 \leqslant i \leqslant n$. Si λ_1, ..., λ_n sont n réels vérifiant $\sum_{i=1}^{n} \lambda_i g(e_i) = 0$, alors

$$(\lambda_1 + \lambda_2)e_1 + \lambda_2 e_2 + ... + \lambda_n e_n = 0,$$

et puisque $(e_1, ..., e_n)$ est une base de E, $\lambda_1 + \lambda_2 = 0$, $\lambda_2 = 0$, ..., $\lambda_n = 0$. La famille $(g(e_i))_{1 \leq i \leq n}$ est donc aussi libre donc g est un automorphisme linéaire de E. Ainsi, $x + f(x) = g(f(x)) = f(g(x)) = f(x)$ donc $x = 0$. Impossible puisque $(x, f(x))$ est libre. ■

Propriété 2.16 *On suppose E euclidien. Le centre de $\mathcal{O}(E)$ est $\{-Id, Id\}$.*

Preuve : Clairement $\{-Id, Id\} \subset Z(\mathcal{O}(E))$. Soit f une isométrie de E qui commute avec toutes les autres isométries. Pour D droite vectorielle de E, on note s_D la réflexion d'axe D. On a dans $\mathcal{O}(E)$:

$$s_{f(D)} = f \circ s_D \circ f^{-1} = s_D,$$

donc f laisse toute droite invariante : f est une homothétie. Son rapport est par nécessité 1 ou -1. ■

2.3.3 Equation des classes

On appelle *centralisateur de a*, le sous-groupe C_a de G formé des éléments g vérifiant $gag^{-1} = a$. C_a est formé des éléments qui commutent avec a. On vérifie facilement que

$$C_a = G \ \Leftrightarrow \ a \in Z(G) \ \Leftrightarrow \ \mathcal{O}_a = \{a\}$$

où $\mathcal{O}_a$ désigne de façon générale l'orbite de a sous l'action de conjugaison, c'est-à-dire l'ensemble $\mathcal{O}_a = \{x \in G \,/\, \exists g \in G \quad x = gag^{-1}\}$. L'application $\tau_a : g \mapsto gag^{-1}$ constitue un paramétrage de l'orbite $\mathcal{O}_a$, et pour mesurer le "surparamétrage", on définit sur G une relation d'équivalence en posant

$$g\mathcal{R}_a g' \ \Leftrightarrow \ gag^{-1} = g'ag'^{-1}.$$

On vérifie que $\mathcal{R}_a$ est en fait la relation de congruence modulo C_a. Ainsi, par factorisation par le quotient, τ_a induit une bijection de G/C_a sur $\mathcal{O}_a$ et si G est fini :

$$\circ(G) = \circ(C_a)\,\sharp(\mathcal{O}_a).$$

Parmi les éléments de G, il y a les "solitaires", les éléments à classe de conjugaison ponctuelle. Ce sont les éléments du centre de G. Soit r le nombre d'orbites non réduites à un point et pour $1 \leqslant i \leqslant r$, a_i un représentant de chacune de ses classes. On a l'égalité

$$\circ(G) = \circ(Z(G)) + \sum_{i=1}^{r} [G : C_{a_i}],$$

connue sous le nom *équations des classes*. Pour une présentation limpide de ces concepts utilisant le langage des actions de groupes, voir [1].

2.3.4 Groupes-quotients

Voici quelques résultats généraux : les preuves manquantes se trouvent par exemple dans [4].

▶ **Théorème d'isomorphisme** :

1) Les relations d'équivalence sur un groupe G compatibles avec la structure de groupe sont les relations modulo un sous groupe distingué de G.

2) Si H est un sous-groupe distingué de G, alors G/H muni de la loi de composition interne "$xH\,.\,yH = xyH$" est un groupe d'élément neutre H. L'application $\Pi : x \mapsto \overline{x} = xH$ est un morphisme surjectif de G sur G/H, de noyau H.

3) Premier théorème d'isomorphisme : Si $f : G \to G'$ est un morphisme de groupes, alors $ker(f)$ est distingué dans G et le groupe quotient $G/ker(f)$ est isomorphe à $Im(f)$. En particulier, $\mathcal{S}_n/\mathcal{A}_n \equiv \mathbb{Z}/2\mathbb{Z}$.

4) On note $Aut(G)$ le groupe des automorphismes de G et $Int(G)$ le sous-groupe des automorphismes intérieurs de G. Montrer que $G/Z(G) \equiv Int(G)$. On pourra considérer le morphisme $Int : a \mapsto [\sigma_a : g \mapsto gag^{-1}]$ de G dans $Aut(g)$.

Remarque : Si G est un groupe d'ordre pair et si H est un sous-groupe de G d'indice 2, alors H est distingué dans G (exercice 2.12) et contient tous les carrés de G. En effet, on a $G/H \equiv \mathbb{Z}/2\mathbb{Z}$ et $\forall x \in G \quad \bar{x}^2 = \overline{x^2} = \bar{1}$ donc $x^2 \in H$. En particulier, si $\mathcal{A}_4$ d'ordre 12 possède un sous-groupe H d'ordre 6, alors H qui contient tous les carrés, contient tous les 3-cycles ($c^3 = 1 \Rightarrow c = (c^{-1})^2$) au nombre de $2 \times C_4^3 = 8$. Impossible !

▶ **Groupe dérivé** :

Dans un groupe G, un élément g est dit *commutateur* s'il existe x et y dans G tels que $g = xyx^{-1}y^{-1}$. Le groupe engendré par les commutateurs est appelé *groupe dérivé* de G et noté $D(G)$.

1) Pour f morphisme de G dans un autre groupe,

$$Im(f) \text{ est commutatif si, et seulement si, } D(G) \subset ker(f).$$

2) Un sous-groupe H de G qui contient $D(G)$ est distingué et tel que G/H est abélien.
En effet, pour $x \in G$ et $h \in H$,

$$xhx^{-1} = xh(x^{-1}h^{-1}hx)x^{-1} = (xhx^{-1}h^{-1})(hxx^{-1}) = (xhx^{-1}h^{-1})\ h$$

est bien dans H. On envisage alors la surjection canonique $p : x \mapsto \overline{x}$ de noyau $ker(p) = H$. Puisque $ker(p)$ contient $D(G)$, $Im(p) = G/H$ est commutatif.

3) $D(G)$ est le plus petit sous-groupe distingué de G (pour l'inclusion) dont le quotient est commutatif.

2.4 A propos de $\mathcal{A}_5$

2.4.1 Centres de $\mathcal{S}_n$ et de $\mathcal{A}_n$

Propriété 2.17 *Pour $n \geqslant 3$, le centre de $\mathcal{S}_n$ est trivial.*

Preuve : Soit s dans $\mathcal{S}_n$ (avec $n \geqslant 3$), $s \neq Id$, i tel que $s(i) = j \neq i$. Soit, à bon droit ($n \geqslant 3$), k différent de i et j, et τ la transposition (j,k). On a : $s\tau(i) = s(i) = j$ et $\tau s(i) = \tau(j) = k$ donc $s \notin Z(\mathcal{S}_n)$. ■

Propriété 2.18 *Pour $n \geqslant 4$, le centre de $\mathcal{A}_n$ est trivial.*

Preuve : Soient $\sigma \in \mathcal{A}_n$, $\sigma \neq Id$, i tel que $\sigma(i) = j \neq i$, k tel que $k \neq i,j$, soient l tel que $l \neq i,j,k$ et enfin définissons $\tau = (j,k)(k,l) \in \mathcal{A}_n$. Les égalités $\sigma\tau(i) = \sigma(i) = j$ et $\tau\sigma(i) = \tau(j) = k$ montrent que $\sigma \notin Z(\mathcal{A}_n)$. ■

2.4.2 Groupe dérivé de $\mathcal{A}_5$

On veut déterminer le groupe dérivé de $\mathcal{A}_5$.

Lemme 2.4 *Les 3-cycles sont conjugués dans $\mathcal{A}_5$ (dans $\mathcal{A}_n$ pour $n \geqslant 5$).*

Preuve : Soit $\tau = (a,b,c)$ et $\tau' = (a',b',c')$ deux 3-cycles. Soit $\sigma \in \mathcal{S}_5$ telle que $a \mapsto a'$, $b \mapsto b'$ et $c \mapsto c'$. Si $\sigma \in \mathcal{A}_5$, c'est fini puisque $\sigma\tau\sigma^{-1} = \tau'$. Sinon, en écrivant $\{1,...,5\} = \{a,b,c,d,e\}$, on pose $\sigma' = (d,e)\sigma \in \mathcal{A}_5$ et on a $\sigma'\tau\sigma'^{-1} = \tau'$. ■

Lemme 2.5 *Pour $n \geqslant 3$, les 3-cycles engendrent $\mathcal{A}_n$.*

Preuve : On sait que $\mathcal{A}_n$ est engendré par les produits de deux transpositions. Or, avec des notations évidentes,

$$(a,b)(b,c) = (b,a)(b,c) = (b,a)(c,b) = (a,b,c),$$

et $(a,b)(c,d) = (a,c,b)(a,c,d)$, donc les 3-cycles sont dans $\mathcal{A}_n$ et engendrent bien $\mathcal{A}_n$. ■

Propriété 2.19 *Pour $n \geqslant 5$, on a : $D(\mathcal{A}_n) = \mathcal{A}_n$.*

Preuve : Un 3-cycle σ et son carré (qui est aussi un 3-cycle) sont conjugués dans $\mathcal{A}_n$ (Lemme 2.4), donc $\sigma^2 = \tau^{-1}\sigma\tau$ avec $\tau \in \mathcal{A}_n$. Par suite le 3-cycle $\sigma = \sigma^{-1}\tau^{-1}\sigma\tau$ est un commutateur. Puisque les 3-cycles engendrent $\mathcal{A}_n$ (Lemme 2.5), on a bien $D(\mathcal{A}_n) = \mathcal{A}_n$. ■

2.4.3 Simplicité de $\mathcal{A}_5$

La taille des orbites (pour la relation de conjugaison), comme la grosseur du centre et du groupe dérivé, mesurent le défaut de commutativité d'un groupe. Par exemple, si G est commutatif, les classes de conjugaison sont des singletons (au nombre de $\circ(G)$). Dans un groupe "fortement non commutatif", les sous-groupes distingués sont à chercher parmi les gros sous-groupes de G, restriction qui laisse penser que les sous-groupes distingués y sont peu nombreux. Aussi, malgré l'exemple des groupes cycliques d'ordre premier, on peut s'attendre à trouver des groupes simples parmi les groupes fortement non commutatifs.

On aura besoin de ces deux résultats :

Lemme 2.6 *Les produits de deux transpositions à supports disjoints sont conjugués dans $\mathcal{A}_5$.*

Preuve : Soit $\tau = (a, b)(c, d)(e)$ et $\tau' = (a', b')(c', d')(e')$. Soit σ dans $\mathcal{S}_5$ envoyant a sur a', b sur b', ..., e sur e', et $\sigma' = (c', d')\sigma$.
On a $\sigma\tau\sigma^{-1} = \sigma'\tau\sigma'^{-1} = \tau'$ avec $\sigma \in \mathcal{A}_5$ ou $\sigma' \in \mathcal{A}_5$. D'où le résultat. ■

Lemme 2.7 *Si $\tau = (a, b, c, d, e)$ et $\tau' = (a', b', c', d', e')$ sont des 5-cycles, alors τ' est conjugué avec τ ou $\tau^2 = (a, c, e, b, d)$ dans $\mathcal{A}_5$.*

Preuve : Soit

$$\sigma = \begin{pmatrix} a & b & c & d & e \\ a' & b' & c' & d' & e' \end{pmatrix}$$

dans $\mathcal{S}_5$, qui vérifie $\tau' = \sigma\tau\sigma^{-1}$. Si σ est une permutation paire, alors τ' est conjuguée à τ dans $\mathcal{A}_5$ comme souhaité. Dans le cas contraire, on envisage le 4-cycle $\gamma = (b, d, e, c) = (b, c)(b, e)(b, d)$, permutation de signature -1 qui conjugue τ et τ^2 dans $\mathcal{S}_5$: $\tau = \gamma\tau^2\gamma^{-1}$. De là, on écrit

$$\tau' = \sigma\tau\sigma^{-1} = \sigma(\gamma\tau^2\gamma^{-1})\sigma^{-1} = (\sigma\gamma)\tau^2(\sigma\gamma)^{-1}$$

où la permutation $\sigma\gamma$ est dans $\mathcal{A}_5$. ■

Propriété 2.20 *Le groupe $\mathcal{A}_5$ est simple.*

Preuve : • **Première preuve (Vive Lagrange !)**
On sait que $\mathcal{A}_5$ a $5!/2 = 60$ éléments. Les 3-cycles et 5-cycles de $\mathcal{S}_5$ sont pairs puisque ce sont des carrés ($c^3 = 1 \Rightarrow c = (c^{-1})^2$ et $c^5 = 1 \Rightarrow c = (c^{-2})^2$).
Dans $\mathcal{A}_5$, il y a :

- le neutre,

- les produits de deux transpositions disjointes (d'ordre 2), au nombre de $\frac{C_5^2 C_3^2}{2} = 15$ (on divise par 2 car $(a,b)(c,d) = (c,d)(a,b)$),

- les 3-cycles au nombre de $C_5^3 \times 2 = 20$ (pour un même support, on a deux 3-cycles distincts),

- les 5-cycles au nombre de $4 \times 3 \times 2 \times 1 = 24$.

On a listé 60 éléments, on a donc toutes les permutations de $\mathcal{A}_5$.

Soit H un sous-groupe distingué de $\mathcal{A}_5$, distinct de $\{1\}$. Si H contient un 3-cycle (respectivement un élément d'ordre 2), il contient sa classe de conjugaison et donc tous les 3-cycles (respectivement tous les produits de deux transpositions disjointes) (Lemmes 2.4 et 2.6). Si H contient un 5-cycle τ, il contient aussi τ^2, et contient donc n'importe quel 5-cycle τ' (Lemme 2.7). Ainsi H contient le neutre et au moins deux permutations d'ordres différents. Sinon, $\circ(H) = 1+24 = 25$, ou $\circ(H) = 1+20 = 21$, ou $\circ(H) = 1+15 = 16$. Impossible puisque 25, 21 et 16 ne divisent pas 60. Il suit : $\circ(H) \geqslant 1 + 15 + 20 = 36$ et toujours avec Lagrange, nécessairement $\circ(H) = 60$, et $H = \mathcal{A}_5$.

• **Seconde preuve (Vive les tricycles !)**
Soit H un sous-groupe distingué de $\mathcal{A}_5$, distinct de $\{1\}$. Si H contient un 3-cycle, il contient tous les 3-cycles (Lemme 2.4), donc $H = \mathcal{A}_5$. Si H contient un 5-cycle $\tau = (a,b,c,d,e)$, il contient le commutateur $(a,b,c)\tau(a,b,c)^{-1}\tau^{-1}$ qui est le 3-cycle (a,b,d). Le sous-groupe H contient alors tous les 3-cycles et $H = \mathcal{A}_5$. Enfin, si H contient un produit de deux transpositions disjointes $\tau = (a,b)(c,d)$, il contient le commutateur

$$\tau(a,b,e)\tau^{-1}(a,b,e)^{-1} = (a,b)(c,d)(a,b,e)(a,b)(c,d)(e,b,a) = (a,b,d)$$

donc $H = \mathcal{A}_5$. ∎

Remarque : $\mathcal{A}_5$ nous fournit un exemple de groupe ne possédant pas de sous-groupe d'ordre égal à un diviseur de l'ordre du groupe : $\mathcal{A}_5$ simple n'a pas de sous-groupe d'ordre 30.

2.4.4 Sous-groupes distingués de $\mathcal{S}_5$

Si $H \neq \{1\}$ est un sous-groupe distingué de $\mathcal{S}_5$, $H \cap \mathcal{A}_5$ est distingué dans $\mathcal{A}_5$ (importance de $\mathcal{A}_5 \triangleleft \mathcal{S}_5$), et puisque $\mathcal{A}_5$ est simple, $H \cap \mathcal{A}_5 = \mathcal{A}_5$ ou $\{1\}$. Si

$H \cap \mathcal{A}_5$ est trivial, la restriction de la signature à H est injective, et même surjective puisque $H \neq \{1\}$, donc H a 2 éléments 1 et σ, où $\circ(\sigma) = 2$. Avec $H \triangleleft \mathcal{S}_5$, on a nécessairement $\tau \circ \sigma \circ \tau^{-1} = \sigma$ pour $\tau \in \mathcal{S}_5$, donc σ est dans le centre de $\mathcal{S}_5$: $\sigma = 1$. Contradiction ! Ainsi, $H \cap \mathcal{A}_5 = \mathcal{A}_5$, $\mathcal{A}_5 \subset H$. Premier cas : $H = \mathcal{S}_5$. Sinon, par Lagrange, $\circ(H) \leqslant n/2$, et par nécessité, $\circ(H) = n/2$ et $H = \mathcal{A}_5$.

Propriété 2.21 *Les seuls sous-groupes distingués de $\mathcal{S}_5$ sont $\{1\}$, $\mathcal{A}_5$ et $\mathcal{S}_5$.*

On admet que

Propriété 2.22 *Pour tout $n \geqslant 5$:*
1. *$\mathcal{A}_n$ est simple.*
2. *Les seuls sous-groupes distingués de $\mathcal{S}_n$ sont $\{1\}$, $\mathcal{A}_n$ et $\mathcal{S}_n$.*

Le groupe alterné $\mathcal{A}_n$ est un sous-groupe de $\mathcal{S}_n$ d'indice 2 (en fait, c'est le seul, cf propriete 2.36 page 89), et $\mathcal{S}_n$ possède des sous-groupes d'indice n (par exemple, $H_n = \{\sigma \in \mathcal{S}_n \ ; \ \sigma(n) = n\}$). Voici un résultat surprenant :

Propriété 2.23 *(Saut d'indice dans $\mathcal{S}_n$)*
Pour $n \geqslant 5$, si H est un sous-groupe de $\mathcal{S}_n$ d'indice $[\mathcal{S}_n : H] = N > 2$, alors $[\mathcal{S}_n : H] = N \geqslant n$.

Preuve : L'idée est de contruire une injection entre $\mathcal{S}_n$ (d'ordre $n!$) et le groupe des permutations de $Q = \mathcal{S}_n/H$ (de cardinal $N!$). On envisage le morphisme $\Phi : s \mapsto [\sigma H \mapsto s\sigma H]$ de $\mathcal{S}_n$ vers $\mathcal{S}_Q$ (action de $\mathcal{S}_n$ sur Q). On choisit à bon droit 3 éléments distincts de Q : H, $\sigma_1 H$ et $\sigma_2 H$. Clairement, les 3 permutations Id, $\Phi(\sigma_1)$, $\Phi(\sigma_2)$ de $\mathcal{S}_Q$ sont distinctes donc $\sharp(Im(\Phi)) \geqslant 3$. Ainsi, $Ker(\Phi)$ est un sous-groupe distingué de $\mathcal{S}_n$ d'indice supérieur à 3, donc $Ker(\Phi) = \{1\}$ et Φ est injective. D'où $n! \leqslant N!$, $n \leqslant N$. ∎

2.5 Carrés non nuls du corps $\mathbb{Z}/p\mathbb{Z}$

Soit $p > 2$ un nombre premier. Certains résultats algébriques valables dans $\mathbb{R}$ ou $\mathbb{C}$ sont corrects dans le corps $\mathbb{Z}/p\mathbb{Z}$. On peut par exemple faire de l'algèbre linéaire et étudier les systèmes $AX = B$ où $A \in GL(n, \mathbb{Z}/p\mathbb{Z})$ et $B \in (\mathbb{Z}/p\mathbb{Z})^n$. On peut aussi s'intéresser à des équations non linéaires, par exemple celles du second degré. Les formules usuelles avec discriminant, sont encore d'actualité et on est alors amené à considérer les éléments de $\mathbb{Z}/p\mathbb{Z}$ qui sont des carrés. On pose : $K = \{x^2 \ ; \ x \in (\mathbb{Z}/p\mathbb{Z})^*\}$.

2.5.1 Morphisme de $(\mathbb{Z}/p\mathbb{Z})^*$ sur $\{-1,1\}$

Si $\varphi : (\mathbb{Z}/p\mathbb{Z})^* \to \{-1,1\}$ est un morphisme de groupes non trivial, φ est alors surjective et son noyau est un sous-groupe de $(\mathbb{Z}/p\mathbb{Z})^*$ d'indice 2. Or le groupe $(\mathbb{Z}/p\mathbb{Z})^*$ est cyclique et d'ordre pair, donc (Prop. 2.6 p. 52) $(\mathbb{Z}/p\mathbb{Z})^*$ ne possède qu'un seul sous-groupe K d'ordre $\frac{p-1}{2}$. On écrit alors $(\mathbb{Z}/p\mathbb{Z})^* = K \cup xK$ où x a été choisi en dehors de K et φ envoie nécessairement tout y de K sur 1 et tout $y \in xK$ sur -1. On a montré qu'il existe au plus un morphisme non trivial de $(\mathbb{Z}/p\mathbb{Z})^*$ sur $\{-1,1\}$.

2.5.2 Un paramétrage de K

On considère l'application $f : x \mapsto x^2$ de $(\mathbb{Z}/p\mathbb{Z})^*$ dans lui-même. On vérifie que f est un morphisme de groupes (pour la loi multiplicative), d'image K et de noyau $ker(f) = \{-1,1\}$ (avec $1 \neq -1$ puisque $p > 2$). Ainsi :

Propriété 2.24 *Dans $(\mathbb{Z}/p\mathbb{Z})^*$, il y a autant de carrés que de non-carrés : autrement dit K possède $(p-1)/2$ éléments.*

Exercice 2.15 *Soient $p > 2$ premier, a et b dans $(\mathbb{Z}/p\mathbb{Z})^*$. Alors il existe x et y dans $\mathbb{Z}/p\mathbb{Z}$ tels que $ax^2 + by^2 = 1$.*

Solution : Dans $\mathbb{Z}/p\mathbb{Z}$, il y a $1 + \frac{p-1}{2} = \frac{p+1}{2}$ carrés (on n'oublie pas 0 !), et par intégrité du corps $\mathbb{Z}/p\mathbb{Z}$, les ensembles $X = \{ax^2 \ ; \ x \in \mathbb{Z}/p\mathbb{Z}\}$ et $Y = \{1 - by^2 \ ; \ y \in \mathbb{Z}/p\mathbb{Z}\}$ ont aussi $\frac{p+1}{2}$ éléments, ce qui assure $X \cap Y \neq \varnothing$.

2.5.3 Comment reconnaître les carrés ?

Pour $x \in (\mathbb{Z}/p\mathbb{Z})^*$, on pose $\phi(x) = x^{\frac{p-1}{2}}$. Par le petit théorème de Fermat ou théorème de Lagrange dans $((\mathbb{Z}/p\mathbb{Z})^*, \times)$, le morphisme de groupes ϕ "tue" les éléments de K. Y a-t-il d'autres éléments dans $ker(\phi)$? Le polynôme $X^{\frac{p-1}{2}} - 1$ à coefficients dans le corps $\mathbb{Z}/p\mathbb{Z}$ a au plus $\frac{p-1}{2}$ racines distinctes, d'où la caractérisation des carrés de K :

Propriété 2.25

1. *Pour $x \in (\mathbb{Z}/p\mathbb{Z})^*$, $x \in K$ si, et seulement si, $x^{\frac{p-1}{2}} = 1$.*
2. *L'application $\left(\frac{\bullet}{p}\right) : x \mapsto x^{\frac{p-1}{2}}$, appelée symbole de Legendre, est le seul morphisme non trivial de $(\mathbb{Z}/p\mathbb{Z})^*$ sur $\{-1,1\}$.*

En particulier,

Propriété 2.26 -1 *est un carré si, et seulement si, $p \equiv 1$ (4).*

Remarque : Une preuve élégante de ce résultat est donné dans la RMS de mars 1986 : on envisage sur $(\mathbb{Z}/p\mathbb{Z})^*$ la relation d'équivalence

$$x\Re y \quad \Leftrightarrow \quad y \in \{x, -x, x^{-1}, -x^{-1}\},$$

on montre qu'il y a une seule classe à deux éléments (si -1 n'est pas un carré de $\mathbb{Z}/p\mathbb{Z}^*$) ou deux classes à deux éléments (si -1 est un carré), les autres classes au nombre de k ayant 4 éléments. Cette partition de $(\mathbb{Z}/p\mathbb{Z})^*$ donne alors : $p-1 = 2+4k$ ou $p-1 = 2+2+4k$...

2.5.4 Symbole de Zolotareff

Soit p un nombre premier impair. Pour n entier premier avec p, on note $\Pi_{n,p}$ la multiplication par n modulo p (endomorphisme du groupe additif $\mathbb{Z}/p\mathbb{Z}$). Puisque le corps $\mathbb{Z}/p\mathbb{Z}$ est intègre, le morphisme $\Pi_{n,p}$ est injectif, donc est une permutation de $\mathbb{Z}/p\mathbb{Z}$. Au passage, on a $(p-1)! = n \times (2n) \times ... \times ((p-1)n)$ d'où : $(n^{p-1}-1)(p-1)! = 0(p)$, et on retrouve :

$$n^{p-1} = 1(p) \quad (\textit{petit théorème de Fermat}).$$

On définit le symbole de Zolotareff $(n \mid p)$ comme étant la signature de $\Pi_{n,p}$. On a, par exemple, $(-1 \mid p) = (-1)^{\frac{p-1}{2}}$ (comme on le voit en décomposant $\Pi_{-1,p} = (1, p-1)(2, p-2)...(\frac{p-1}{2}, \frac{p+1}{2})$, d'où l'équivalence

$$(-1 \mid p) = 1 \ \Leftrightarrow \ p \equiv 1(4).$$

Par ailleurs, pour $(n,m) \in \mathbb{Z}^2$, $n \wedge p = 1$, $m \wedge p = 1$, on a : $nm \wedge p = 1$ et $\Pi_{nm,p} = \Pi_{n,p} \circ \Pi_{m,p}$ donc par le morphisme signature, le symbole de Zolotareff est multiplicatif.

Propriété 2.27 *(Lemme de Zolotareff)*
Soit $p > 2$ premier et n un entier premier à p. La signature $(n \mid p)$ de la multiplication $\Pi_{n,p}$ par n, est égale au symbole de Legendre $\left(\frac{n}{p}\right)$.

Preuve : Soit x un générateur de $(\mathbb{Z}/p\mathbb{Z})^*$. On écrit $n = x^k$ avec $k \in \mathbb{Z}$ et la multiplication par n est l'itérée $(\Pi_{x,p})^k$. Que vaut la signature $\varepsilon(\Pi_{x,p})$?
Si on écrit $\mathbb{Z}/p\mathbb{Z} = \{0, x, x^2, ..., x^{p-1} = 1\}$, la multiplication par x est un $(p-1)$-cycle, donc est impaire. Ainsi $\varepsilon(\Pi_{n,p}) = (-1)^k$. Par ailleurs, le symbole de Legendre $\left(\frac{n}{p}\right)$ est : $n^{\frac{p-1}{2}} = (x^{\frac{p-1}{2}})^k$. Or x n'est pas un carré modulo p (sinon, tous les éléments de $\mathbb{Z}/p\mathbb{Z}$ seraient des carrés), donc $\left(\frac{n}{p}\right) = (-1)^k$. D'où le résultat. ■

2.6 Groupes d'ordre p^2

2.6.1 Détermination des groupes d'ordre 4

▶ **Les 2 modèles :**

Soit G un groupe d'ordre 4 non cyclique. L'ordre de tout élément distinct de 1 est un diviseur de 4, autre que 1 et 4 : c'est 2. On peut donc écrire $G = \{1, a, b, c\}$ où $a^2 = b^2 = c^2 = 1$. L'élément ab est clairement distinct de 1, a, b, ce qui impose $ab = c$. On fabrique ainsi sans problème la table de composition de G.
Par ailleurs, on construit facilement un isomorphisme ϕ de G sur le groupe additif $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ en posant $\phi(1) = (0, 0)$, $\phi(a) = (1, 0)$, $\phi(b) = (0, 1)$, $\phi(c) = (1, 1)$. G est appelé *groupe de Klein* et noté V_4 (Vierergruppe).

▶ **Une réalisation de** V_4 :

Dans le groupe $\mathcal{A}_4$ (groupe alterné de degré 4), les produits de 2 transpositions disjointes (qui sont d'ordre 2)

$$u = (1, 2)\,(3, 4), \quad v = (1, 3)\,(2, 4), \quad w = (1, 4)\,(2, 3)$$

et Id forment un sous-groupe dont aucun élément n'est d'ordre 4 : c'est V_4.
On peut noter que, pour a, b, c, d distincts dans $\{1, 2, 3, 4\}$ et pour σ dans $\mathcal{A}_4$ (resp $\mathcal{S}_4$),

$$\sigma\,(a, b)\,(c, d)\,\sigma^{-1} = \sigma\,(a, b)\,\sigma^{-1}\sigma\,(c, d)\,\sigma^{-1} = (\sigma(a), \sigma(b))\,(\sigma(c), \sigma(d)).$$

Ceci prouve que V_4 est distingué dans $\mathcal{A}_4$ (resp dans $\mathcal{S}_4$) et que $\mathcal{A}_4$ n'est pas simple.

Remarque : On illustre une nouvelle fois la non-transitivité de $\lhd$. Le sous-groupe $< u = (1, 2)(3, 4) >$ est distingué dans V_4, lui-même distingué dans $\mathcal{A}_4$, et pourtant $< (1, 2)(3, 4) >$ n'est pas distingué dans $\mathcal{A}_4$. En effet :

$$(1, 3)\ (1, 2)(3, 4)\ (1, 3) = (1, 4)(2, 3) \notin < (1, 2)(3, 4) > .$$

Puisque $\{u, v, w\}$ est une classe de conjugaison dans $\mathcal{S}_4$, on définit à bon droit l'application

$$\Phi\ :\ \sigma \mapsto \begin{cases} \{u, v, w\} & \to \{u, v, w\} \\ z & \mapsto \sigma z \sigma^{-1} \end{cases},$$

(action de $\mathcal{S}_4$ sur $\{u, v, w\}$ par conjugaison) qui est morphisme de groupes de $\mathcal{S}_4$ dans $\mathcal{S}_{\{u,v,w\}} \equiv \mathcal{S}_3$. On vérifie que chaque transposition de $\mathcal{S}_{\{u,v,w\}}$

est atteinte, par exemple $(v, w) = \Phi((1, 2))$. Ainsi, Φ est surjective (puisque $\mathcal{S}_{\{u,v,w\}}$ est engendré par ses transpositions), ce qui donne :

$$\sharp(ker(\Phi)) = \frac{4!}{3!} = 4.$$

Or, $V_4 = \{Id, u, v, w\} \subset ker(\Phi)$, d'où :

$$ker(\Phi) = V_4 \quad \text{et} \quad \mathcal{S}_4/V_4 \equiv \mathcal{S}_3.$$

Interprétation (M. Alessandri) : "il y a 3 façons de grouper 4 objets distincts deux par deux. Une permutation de ces 4 objets induit une permutation des 3 facons de les grouper."

► **Une réalisation géométrique de V_4 :**

Dans le plan affine euclidien, $ABCD$ désigne un rectangle non carré de centre O. On cherche l'ensemble $\mathcal{I}$ (en fait le groupe) des isométries affines conservant globalement $ABCD$. On admet que les éléments de $\mathcal{I}$ sont exactement les isométries conservant les 4 sommets A, B, C et D. Soit $f \in \mathcal{I}$. L'application affine f conserve le barycentre donc $f(O) = O$. Par conservation des longueurs, l'image de $[AB]$ est soit $[AB]$, soit $[CD]$.

- Si $f([AB]) = [AB]$, alors le milieu I de $[AB]$ est fixé. Ainsi la droite (OI) est fixée : $f = Id$ ou f est la réflexion s d'axe (OI).

- Si $f([AB]) = [CD]$, alors A est envoyé sur C ou D.

- Si $A \mapsto C$, en notant s_O la symétrie centrale de centre O, $s_O \circ f$ fixe A et O, donc $s_O \circ f = Id$, ou $s_O \circ f$ est la réflexion d'axe (AC). Ce dernier cas est impossible : B serait envoyé sur D, les diagonales (AC) et (BD) seraient perpendiculaires, ce qui est exclu puisque $ABCD$ est supposé non carré. Ainsi $f = s_O$.

- Si $A \mapsto D$, puisque $AD = Df(D)$, $f(D)$ est nécessairement A. Ainsi le milieu J de $[AD]$ est fixé. On a alors $f = Id$ ou f est la réfléxion d'axe (OJ). On vérifie réciproquement que les involutions Id, s_O, $s_{(OI)}$ et $s_{(OJ)}$ sont dans $\mathcal{I}$ d'où :

$$\mathcal{I} = \{Id, s_O, s_{(OI)}, s_{(OJ)}\} \equiv V_4.$$

► **Sous-groupes du groupe quaternionique $\mathcal{H}_8$:**

On peut montrer que dans $GL(2, \mathbb{C})$, les matrices

$$I_2,\ -I_2,\ a = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix},\ -a,\ b = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},\ -b,\ c = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix},\ \text{et}\ -c$$

forment un groupe noté $\mathcal{H}_8$ et appelé *groupe quaternionique*. Soit H un sous-groupe de $\mathcal{H}_8$. Si $H \neq \{1\}$ et $H \neq \mathcal{H}_8$, H possède 2 ou 4 éléments.

- Si $H = \{I_2, \gamma\} \equiv \mathbb{Z}/2\mathbb{Z}$, alors $\gamma \neq I_2$ vérifie $\gamma^2 = I_2$. Or le seul élément de H d'ordre 2 est $-I_2$. Ainsi $H = \{I_2, -I_2\}$ et

$$\forall g \in \mathcal{H}_8 \quad \forall h \in H \quad ghg^{-1} = h \in H.$$

- Si H est d'ordre 4 (nécessairement isomorphe à $\mathbb{Z}/4\mathbb{Z}$ puisque le seul élément d'ordre 2 dans $\mathcal{H}_8$ est $-I_2$), il est d'indice 2 dans $\mathcal{H}_8$, et donc distingué dans $\mathcal{H}_8$. Le sous-groupe $< a >$ est un exemple de sous-groupe de $\mathcal{H}_8$ d'ordre 4.

Bilan : Le groupe $\mathcal{H}_8$ est non commutatif et pourtant tous ses sous-groupes sont distingués. Cela laisse penser que $\mathcal{H}_8$ est un groupe des plus commutatifs parmi les groupes non commutatifs. Voici aussi un exemple de groupe non cyclique dont tous les sous-groupes propres sont cycliques.

Exercice 2.16 *Montrer que* $D(\mathcal{H}_8) = \{I_2, -I_2\}$. *Reconnaître le quotient* $\mathcal{H}_8/D(\mathcal{H}_8)$.

Solution : $\mathcal{H}_8/\{I_2, -I_2\}$, qui est d'ordre 4, est commutatif (isomorphe à $\mathbb{Z}/4\mathbb{Z}$ ou V_4) donc $D(\mathcal{H}_8) \subset \{I_2, -I_2\}$. Ainsi, $D(\mathcal{H}_8) = \{I_2\}$ ou $\{I_2, -I_2\}$. Or $D(\mathcal{H}_8) \neq \{I_2\}$ (sinon $\mathcal{H}_8$ serait commutatif), donc

$$D(\mathcal{H}_8) = \{I_2, -I_2\}.$$

Pour tout $g \in \mathcal{H}_8$, $g^2 = I_2$ ou $g^2 = -I_2$, donc tout élément de $\mathcal{H}_8/\{I_2, -I_2\}$ est involutif. Ainsi, $\mathcal{H}_8/\{I_2, -I_2\}$, qui est d'ordre 4, est isomorphe à V_4.

Remarque : Pour tout sous-groupe H de $\mathcal{H}_8$, $H \cap \{I_2, -I_2\} \neq \{I_2\}$. On dit que le sous-groupe $\{I_2, -I_2\}$ est *dense* dans $\mathcal{H}_8$. On note également que tout sous-groupe H contient $D(\mathcal{H}_8)$ et on retrouve ainsi que H est distingué dans $\mathcal{H}_8$.

Exercice 2.17 *Soit G un groupe fini et D un sous-groupe de G contenant tous les éléments de G d'ordre premier. Montrer que D est dense dans G.*

Solution : Soit H un sous-groupe de G non trivial. Soit $h \in H$, $h \neq 1$. On suppose $h \notin D$. On appelle δ l'ordre de h, qui n'est pas premier. On choisit alors un diviseur d premier de δ, et un $g \in < h >$ d'ordre d. L'élément g est dans H et dans D (par hypothèse).

2.6.2 Détermination des groupes d'ordre 9

Soit G un groupe d'ordre 9. Si G n'est pas cyclique, tout élément x de $G\setminus\{1\}$ est d'ordre 3. Soit $a \in G\setminus\{1\}$ et $b \in G\setminus < a >$. On a $a^2 = a^{-1} \neq b$ et $b^2 = b^{-1} \neq a$, donc le sous-groupe $< a, b >$ engendré par a et b contient les 9 éléments distincts $\{1, a, a^2, b, b^2, ab, ab^2, a^2b, a^2b^2\}$. Ainsi $G =< a, b >$. L'élément ba, clairement distinct de 1, a, a^2, et b^2, appartient donc à $\{ab, a^2b, ab^2, a^2b^2\}$.

- Si $ba = a^2b^2$, alors $(ba)^2 = baba = a^2b^2ba = a^2b^3a = a^3 = 1$, donc $\circ(ba) = 2$. Impossible puisque 2 ne divise pas 9.

- Si $ba = ab^2$, alors $(ba)^2 = baba = ab^3a = a^2$. Il vient alors $(ba)^3 = b^2$, $(ba)^5 = a^2b^2$, $(ba)^6 = baa^2b^2 = 1$, ce qui est impossible puisque 6 ne divise pas 9.

- On montre de même que si $ba = a^2b$, alors $(ba)^6 = 1$, ce qui est impossible.

Bilan : les générateurs a et b de G commutent donc G est abélien. On envisage alors l'application

$$\phi : \ < a > \times < b > \to \ G = \{1, a, a^2, b, b^2, ab, ab^2, a^2b, a^2b^2\}.$$

ϕ est un morphisme surjectif de groupes commutatifs. Comme $< a > \times < b >$ et G ont même cardinal, ϕ est un isomorphisme de groupe, et l'on peut écrire : $\mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z} \simeq G$.

2.6.3 Cas général

Définition 2.1 *Un groupe d'ordre p^k où p est un entier premier et où $k \in \mathbb{N}^*$, est appelé un p-*groupe*.*

Lemme 2.8 *Le centre d'un p-groupe n'est jamais trivial.*

Preuve : Soit G un p-groupe d'ordre p^k. Avec l'équation aux classes, on a $p^k = \circ(Z(G)) + \sum_{x\in T}[G : C_x]$ où T désigne une transversale aux classes de conjugaison non ponctuelles. Or pour $x \in T$, $[G : C_x] > 1$ et $[G : C_x] \mid p^k$, donc $p \mid \circ(Z(G))$, donc $\circ(Z(G)) \geqslant p$. ■

Exemple : Le centre Z de $\mathcal{H}_8$ a 2 ou 4 éléments. Clairement $\{I_2, -I_2\} \subset Z$. Si $\circ(Z) = 4$, on choisit $x \in \mathcal{H}_8 \setminus Z$ et $Z \subsetneq Z \cup \{x\} \subset C_x$ donne $C_x = \mathcal{H}_8$ (avec Lagrange), donc x commute avec tous les éléments de $\mathcal{H}_8$. Absurde puisque $x \notin Z$. Ainsi $\circ(Z) = 2$ et $Z = \{I_2, -I_2\}$.

Exercice 2.18 *Un groupe G d'ordre p^k (p premier, $k \geqslant 2$) n'est pas simple.*

Solution : - Si G est abélien. Soit $x \in G$, $x \neq 1$. Si $< x > \neq G$, $< x >$ convient. Sinon, $< x^p >$ convient.

- Si G n'est pas commutatif, on a $Z(G) \neq G$, $Z(G) \neq \{1\}$ et $Z(G)$ distingué dans G.

Lemme 2.9 *Si p est premier et si G est un groupe d'ordre p^2, alors G est commutatif.*

Preuve : • *Première méthode* : On raisonne par l'absurde. Soit $x \in G \setminus Z(G)$. Le centralisateur C_x contient $Z(G)$ (donc au moins p éléments), et $x \notin Z(G)$, donc $\circ(C_x) \geqslant p+1$, et puisque $\circ(C_x) \mid \circ(G)$, on a $\circ(C_x) = p^2$, $C_x = G$, et par conséquent $x \in Z(G)$. Contradiction.

• *Seconde méthode* : On raisonne par l'absurde en utilisant le lemme 2.10. Si G n'est pas commutatif, alors $Z(G)$ non trivial est d'ordre p. Le groupe $G/Z(G)$ est d'ordre p premier, donc cyclique, donc G est commutatif. Contradiction. ■

Voici le résultat admis utilisé dans la preuve précédente :

Lemme 2.10 *Si $G/Z(G)$ est cyclique, alors G est commutatif.*

Propriété 2.28 *Soit p premier. Il y a exactement deux modèles de groupes d'ordre p^2 : le groupe cyclique $\mathbb{Z}/p^2\mathbb{Z}$ et $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$. Ces deux groupes sont commutatifs.*

Preuve : On suppose G non cyclique. Soit $a \in G$, $a \neq 1$. On a $\circ(a) = p$. Soit à présent $b \in G$ tel que $b \notin < a >$. On remarque qu'on a aussi $\circ(b) = p$. On envisage alors l'application $\phi : < a > \times < b > \to G$ définie par $\phi((x, y)) = xy$. Avec G commutatif (d'après le lemme 2.9), on vérifie facilement que ϕ est un morphisme de groupe. Son image est un sous-groupe de G qui contient $\{1, a, .., a^{p-1}, b\}$, donc plus de $p+1$ éléments. Comme son cardinal vaut p ou p^2, c'est p^2. Ainsi ϕ est surjective et $G \simeq \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$. ■

Remarque : Si G est un groupe d'ordre p^2 et H un sous-groupe propre de G, alors l'ordre de H est p et H est cyclique. $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ est donc un exemple de groupe *commutatif* non-cyclique dont tous les sous-groupes propres sont cycliques.

2.6.4 Réciproque de Lagrange dans les p-groupes

Propriété 2.29 *Soit p un nombre premier, $k \in \mathbb{N}^*$ et G un groupe d'ordre $n = p^k$. Alors, pour tout diviseur p^m ($0 \leqslant m \leqslant k$) de $n = p^k$, il existe un sous-groupe de G d'ordre p^m.*

Preuve : On procède par récurrence sur k. Si $k = 1$, p premier n'admet que les diviseurs 1 et p, et les sous-groupes triviaux de G conviennent. On suppose la propriété vraie pour tous les groupes d'ordre p^k. Soit G un groupe d'ordre p^{k+1} et $0 \leqslant m \leqslant k+1$. Si $m = 0$, c'est fini puisque $\{1\}$ est d'ordre $p^0 = 1$. On suppose $m > 0$. Puisque le centre de G n'est pas trivial (lemme 2.8), on choisit $z \in Z = Z(G)$, $z \neq 1$. Dans le sous-groupe cyclique $< z >$ dont l'ordre est nécessairement une puissance de p, on prend à bon droit h d'ordre p et on note $H = < h >$. Puisque $H \subset Z$, H est distingué dans G et le groupe quotient G/H (d'ordre p^k) possède un sous-groupe K d'ordre p^{m-1}. On pose : $L = \pi^{-1}(K)$ où π désigne la surjection $x \mapsto \overline{x} = xH$. L est un sous-groupe de G, et la restriction de π à L est un morphisme surjectif de L sur son image $\pi(L) = K$, donc $L/ker(\pi_{|L}) = L/H \equiv K$, en particulier : $\circ(L) = \circ(H) \times \circ(K) = p^m$. ■

2.7 Théorème de Dixon

Soit G un groupe fini non commutatif. On tire successivement (avec remise) deux éléments de G, et on s'intéresse à la probabilité $P(G)$ pour que ces éléments commutent (cf [2]).

2.7.1 Dans $\mathcal{S}_3$

On étiquette les éléments de $\mathcal{S}_3$: Id, $\tau_1 = (2,3)$, $\tau_2 = (1,3)$, $\tau_3 = (1,2)$, $r = (1,2,3)$ et $\rho = r^{-1}$. La permutation Id est dans le centre de $\mathcal{S}_3$, τ_i commute avec Id et τ_i tandis que r et ρ commutent avec Id, r et ρ donc la probabilité cherchée est

$$P(\mathcal{S}_3) = \frac{6 + 3 \times 2 + 2 \times 3}{6^2} = \frac{1}{2}.$$

2.7.2 Théorème

Propriété 2.30 *Soit G un groupe fini non commutatif. On tire successivement (avec remise) deux éléments de G. Alors la probabilité $P(G)$ pour que ces éléments commutent vérifie :*

$$P(G) \leqslant \frac{5}{8}.$$

Preuve : • *Remarque préliminaire* : Le centre $Z(G)$ étant un sous-groupe de G, par le théorème de Lagrange, il existe $m \in \mathbb{N}^*$ tel que : $\circ(G) = m \times \circ(Z(G))$. Puisque G est non commutatif, $Z(G) \neq G$ et $m \geqslant 2$.
Pour $x \in G$, Le centralisateur C_x de x est aussi un sous-groupe de G, donc il existe $k_x \in \mathbb{N}^*$ tel que $\circ(G) = k_x \times \circ(C_x)$. Par ailleurs, $Z(G)$ étant un

sous-groupe de C_x, il existe $j_x \in \mathbb{N}^*$ tel que $\circ(C_x) = j_x \times \circ(Z(G))$. On peut écrire :

$$\forall x \in G \quad m = k_x\, j_x.$$

Pour $x \notin Z(G)$, il existe $y \in G$ tel que $xy \neq yx$, donc $C_x \neq G$, donc $k_x \geqslant 2$. De plus $x \in C_x$, donc $C_x \neq Z(G)$, donc $j_x \geqslant 2$. En définitive, $m \geqslant 4$.

• Le nombre de tirages possibles est : $(\circ(G))^2$. Les couples favorables sont dans $\bigcup_{x \in G} \{(x, y) \in G^2 \ ; \ y \in C_x\}$. La probabilité cherchée est donc :

$$P(G) = \frac{1}{(\circ(G))^2} \sum_{x \in G} \circ(C_x).$$

On a : $P(G) = \frac{1}{(\circ(G))^2}(\sum_{x \in Z(G)} + \sum_{x \notin Z(G)})$, c'est-à-dire

$$P(G) = \frac{1}{(\circ(G))^2} \left\{ \circ(Z(G)) \circ (G) + \sum_{x \notin Z(G)} \circ(C_x) \right\}.$$

On peut remarquer que pour $x \notin Z(G)$, $k_x \geqslant 2$ donc : $\circ(C_x) \leqslant \dfrac{\circ(G)}{2}$ (✠). Ainsi

$$P(G) \leqslant \frac{1}{(\circ(G))^2} \left\{ \circ(Z(G)) \circ (G) + [\circ(G) - \circ(Z(G))] \times \frac{\circ(G)}{2} \right\},$$

$$P(G) \leqslant \frac{\circ(Z(G))}{\circ(G)} + \frac{1}{2} - \frac{1}{2}\frac{\circ(Z(G))}{\circ(G)} \leqslant \frac{1}{2}\frac{1}{m} + \frac{1}{2}.$$

On rappelle que $\dfrac{1}{m} \leqslant \dfrac{1}{4}$ (♣). Il vient :

$$P(G) \leqslant \frac{1}{8} + \frac{1}{2} = \frac{5}{8}. \quad \blacksquare$$

2.7.3 Constante optimale

Si $P(G) = 5/8$, alors il y a égalité dans l'inégalité (♣), donc

$$m = \frac{\circ(G)}{\circ(Z(G))} = 4.$$

Réciproquement, pour $x \notin Z(G)$, les assertions $k_x j_x = 4$, $k_x \geqslant 2$, et $j_x \geqslant 2$ imposent $j_x = k_x = 2$. Il y a alors égalité dans les inégalités (✠) et (♣), ce qui donne $P(G) = 5/8$. En définitive,

$$P(G) = \frac{5}{8} \quad \Leftrightarrow \quad [G : Z(G)] = 4.$$

On exhibera dans la suite du texte des groupes finis G vérifiant $P(G) = 5/8$.

2.8 Théorème de Cauchy

2.8.1 Groupes d'exposant 2

Propriété 2.31 *Soit G un groupe tel que :* $\forall x \in G \quad x^2 = 1$. *Alors :*

1. *G est commutatif.*
2. *Si G est fini, $\circ(G) = n$ est une puissance de 2 ($n = 2^q$) et $G \simeq (\mathbb{Z}/2\mathbb{Z})^q$.*

Preuve : Pour u et v dans G, on écrit $uv\ uv = 1$, donc $uvu = v$, donc $uv = vu$. Pour la deuxième assertion, on raisonne par récurrence sur l'ordre du groupe.

Initialisation : Si $G = \{1, x\}$, on a bien $2 = 2^1$ et $G \simeq \mathbb{Z}/2\mathbb{Z}$.

Hérédité : On suppose qu'un groupe dont l'ordre est inférieur ou égal à $n-1$ et dans lequel tout élément est involutif, a un cardinal qui est une puissance de 2.

• *Première méthode* : Soient $x_1, .., x_m \in G$ tels que $G =< x_1, .., x_m >$ et $x_m \notin < x_1, ..x_{m-1} >$. On écrit :

$$G = \left\langle \underbrace{< x_1, .., x_{m-1} >}_{H} \cup \{x_m\} \right\rangle .$$

On pose :

$$\mathcal{G} = \{hx_m^k \ ; \ h \in H, k \in \mathbb{Z}\} = \{hx_m^k \ ; \ h \in H, 0 \leqslant k < 2\} = H < x_m > .$$

La partie $\mathcal{G}$ de G est un sous-groupe de G (importance de G abélien), qui contient $H \cup \{x_m\}$, donc $\mathcal{G} = G$. A présent $(h, k) \in H \times \{0, 1\} \mapsto hx_m^k \in G$ est une application surjective, injective (facile à vérifier), donc $n = \circ(G) = 2\circ(H)$. Avec $\circ(H) \leqslant n - 1$ et compte tenu de l'hypothèse de récurrence, n est bien une puissance de 2.
On peut ensuite considérer $(h, x) \in H\times < x_m > \mapsto hx \in G = H \ < x_m >$, qui est un morphisme de groupe (importance de G commutatif), surjectif, injectif (avec $H\cap < x_m >= \{1\}$), ce qui donne $G \simeq (\mathbb{Z}/2\mathbb{Z})^q$.

• *Deuxième méthode* : Soit $x \in G$, $x \neq 1$. Grâce à la commutativité de G, pour tous $a, b, \alpha, \beta \in G$, on vérifie facilement que $\alpha < x >= a < x >$ et $\beta < x >= b < x >$ impliquent $\alpha\beta < x >= ab < x >$. On définit ainsi à bon droit une l.c.i sur $G/ < x >$ en posant $a < x > \ b < x >= ab < x >$. On montre sans problème que $G/ < x >$ muni de cette loi a une structure de groupe. Dans $G/ < x >$ (de cardinal $n/2 < n$), tout élément est clairement involutif et par hypothèse de récurrence, $n/2$ est une puissance de 2. D'où le résultat. ■

Exercice 2.19 *Un groupe dans lequel tout élément distinct du neutre est d'ordre 3, est-il nécessairement commutatif ?*

Exercice 2.20

1) Soit G un groupe. On suppose que le groupe $Aut(G)$ des automorphismes de G est d'ordre $p > 2$ premier. Montrer qu'il existe un entier $n \geq 2$ tel que $G \equiv (\mathbb{Z}/2\mathbb{Z})^n$.

2) En déduire qu'il n'existe pas de tel groupe G.

Solution : 1) Le groupe $Aut(G)$ est cyclique (puisque d'ordre p premier) donc le sous-groupe $Int(G) \equiv G/Z(G)$ des automorphismes intérieurs est lui-même cyclique, donc G est commutatif d'après le lemme 2.10. A présent le passage à l'inverse $i : g \mapsto g^{-1}$ est un morphisme de G dans lui-même (importance de G commutatif), involutif, donc $i \in Aut(G)$. Puisque l'ordre de i ne peut être 2 (qui ne divise pas p), $i = Id_G$.
Ainsi, tout élément de G est d'exposant 2 et $G \equiv (\mathbb{Z}/2\mathbb{Z})^n$ avec $n \geqslant 1$ d'après la propriété 2.31. On termine en remarquant que $\mathbb{Z}/2\mathbb{Z}$ n'a qu'un seul automorphisme, donc nécessairement $n \neq 1$.

2) Le lemme 2.11 ci-dessous montre que $Aut(\mathbb{Z}/2\mathbb{Z})^n)$ n'est pas un groupe d'ordre premier. Il n'existe donc pas de tel groupe G.

Lemme 2.11 *L'ordre β_n de $Aut((\mathbb{Z}/2\mathbb{Z})^n)$ est égal à $\beta_n = \prod_{i=0}^{n-1}(2^n - 2^i)$.*

Preuve : Les automorphismes du groupe $((\mathbb{Z}/2\mathbb{Z})^n, +)$ sont exactement les automorphismes linéaires du $\mathbb{Z}/2\mathbb{Z}$-espace vectoriel $(\mathbb{Z}/2\mathbb{Z})^n$, donc β_n est aussi le cardinal de l'ensemble des bases de $(\mathbb{Z}/2\mathbb{Z})^n$.
En effet, en notant $(e_1, ..., e_n)$ la base canonique de $(\mathbb{Z}/2\mathbb{Z})^n$, on construit une bijection de $GL(n, \mathbb{Z}/2\mathbb{Z})$ sur l'ensemble des bases en envoyant toute matrice A de $GL(n, \mathbb{Z}/2\mathbb{Z})$ sur $(Ae_1, ..., Ae_n)$.
A présent, pour choisir une base $(a_1, ..., a_n)$ de $(\mathbb{Z}/2\mathbb{Z})^n$, on choisit a_1 quelconque non nul, ce qui donne $2^n - 1$ choix pour a_1. Les vecteurs a_1, ..., a_i étant choisis, on prend a_{i+1} en dehors du sous-espace vectoriel $(a_1, ..., a_i)$, ce qui laisse $2^n - 2^i$ choix pour a_{i+1}. Ainsi $\beta_n = \prod_{i=0}^{n-1}(2^n - 2^i)$. ■

2.8.2 Groupes diédraux

Pour $n \geqslant 2$, on pose : $\omega = \exp \frac{2i\pi}{n}$ et pour $0 \leqslant k < n$, on note A_k le point d'affixe ω^k. On veut l'ensemble (en fait le groupe) D_n des isométries affines du plan euclidien conservant globalement le polygone régulier $A_0 A_1..A_{n-1}$. C'est aussi le groupe des isométries conservant l'ensemble $\{A_0, ..., A_{n-1}\}$ des n sommets.

Soit $f \in D_n$. L'isométrie f conserve le barycentre donc $f(O) = O$. Pour l'image de A_0, on a a priori n possibilités (par exemple A_α) et par conservation des longueurs, l'image de A_1 est soit $A_{\alpha-1}$ ou $A_{\alpha+1}$ (les images des autres sommets étant alors parfaitement déterminées).
Ceci donne donc l'information : $\sharp(D_n) \leqslant 2n$.

- Si $f(A_0) = A_0$, alors f laisse invariante la droite (OA_0) donc f est la réflexion s d'axe (OA_0) ou l'identité.

- Si A_0 est envoyé sur un A_i $(1 \leqslant i \leqslant n-1)$, en notant r la rotation $z \mapsto \omega z$ (de centre O et d'angle $\frac{2\pi}{n}$) et $k = n - i$, la rotation r^k vérifie $r^k \circ f(A_0) = A_0$. Ainsi, $r^k \circ f = Id = r^n$ ou $r^k \circ f = s = r^n s$. *Bilan* : $f = r^{n-k} = r^i$ ou $f = r^{n-k}s = r^i s$.

En conclusion :

Propriété 2.32 $\{Id, r, ..., r^{n-1}, s, rs, ..., r^{n-1}s\}$ *est le groupe des isométries conservant un polygone régulier à n sommets $A_0...A_n$. Il est est formé de n rotations et de n réflexions.*

On démontre que :

Lemme 2.12 *Soit $n \geqslant 2$. Deux groupes G et $\mathcal{G}$ vérifiant :*
1) $\circ(G) = \circ(\mathcal{G}) = 2n$,
2) $G = < r, s >$ *et* $\mathcal{G} = < \rho, \sigma >$,
3) $\circ(s) = \circ(\sigma) = 2$ *et* $\circ(r) = \circ(\rho) = n$,
4) $\circ(rs) = \circ(\rho\sigma) = 2$,
sont isomorphes.

Et l'on pose :

Définition 2.2 *Un groupe vérifiant les quatre conditions du lemme précédent est appelé* groupe diédral d'indice n *et noté* D_n.

Remarques : α) $D_2 = V_4$ est commutatif. Pour $n \geqslant 3$, l'égalité $srsr = 1$ permet d'affirmer que D_n n'est pas commutatif (sinon $\circ(r) = n = 2$).

β) On illustre une nouvelle fois la non-transitivité de $\lhd$. On considère pour cela le groupe D_4 des isométries du plan complexe qui conservent le carré de sommets 1, i, -1 et $-i$. On note s la symétrie d'axe $y = 0$ et r la rotation de centre O qui envoie 1 sur i. On envisage alors les sous-groupes $H = < s >$ et $K = < s, -Id >$. Puisque K est commutatif (ses générateurs commutent), on a $H \lhd K$. Par ailleurs, l'égalité $rsr^{-1} = r^2 s = -s \in K \setminus H$ montre que $H \not\lhd D_4$ et $K \lhd D_4$.

Exercice 2.21 *Soit n un entier naturel ≥ 3. Déterminer selon la parité de n, le centre Z et le groupe dérivé D du groupe diédral D_n d'indice n.*

Solution : • On commence par une *remarque préliminaire* : dans le groupe $D_n =< r, s >$, pour $0 \leqslant i \leqslant n$, $r^i s$ est une réflexion, donc $r^i s\, r^i s = 1$, donc $r^i s = s r^{-i}$.

• *Recherche de Z*

- Si n est impair, soit $\phi \in Z$, $\phi \neq 1$. Si ϕ est une rotation, $\phi = r^i$ où $0 \leqslant i < n$. On a alors $\phi s = s\phi$, $r^i s = s r^i$, $r^i = r^{-i}$, $r^{2i} = 1$, $2i = n$, $2 \mid n$. Absurde ! Si ϕ est une réflexion, $\phi = r^i s$, et $r^i s\; s = s\; r^i s$ donne $r^i = r^{-i}$, $2i = n$ et de nouveau une contradiction avec n impair. En définitive, le centre Z de D_n est $\{1\}$ lorsque n est impair.

- Reste à envisager le cas où $n = 2k$ est pair. La rotation r^k commute avec toute rotation r^i, et avec toute réflexion $r^i s$. En effet,

$$r^k\, r^i s = r^i s\, r^k \quad \Leftrightarrow \quad r^{k+i} = r^{i-k} \quad \Leftrightarrow \quad r^{2k} = 1,$$

qui est bien vrai. Une rotation r^i $(0 < i < k)$ ne commute pas avec s : si $r^i s = s r^i$, $r^{2i} = 1$, ce qui n'est pas. Si $r^j \in Z$ avec $k < j < n$, $r^{-j} = r^{n-j} \in Z$ avec $0 < n - j < k$, ce qui est impossible d'après ce qui précède. Clairement, $s \notin Z$ puisque $rs = sr$ conduit à l'impossibilité $r^2 = 1$. La réflexion $r^i s$ $(0 < i < k)$ n'est pas non plus dans Z puisque $r^i s\; s = s\; r^i s$ conduit à $r^{2i} = 1$. De même, $r^j s \notin Z$ lorsque $k < j < n$. Qu'en est-il de $r^k s$? Si $r^k s\; r = r\; r^k s$, $r^{k-1} = r^{k+1}$, $r^2 = 1$, qui est faux. En résumé, le centre Z de D_{2k} est $\{1, r^k\} \equiv \mathbb{Z}/2\mathbb{Z}$.

• *Recherche de D*

Puisque $s^{-1}r^{-1}sr = sr^{-1}sr = sr^{-1}r^{-1}s = sr^{-2}s = r^2$, r^2 est un commutateur donc $< r^2 > \subset D$. Or $< r^2 >$ est d'ordre n si n est impair, $n/2$ si n est pair, donc $D_n/ < r^2 >$ est d'ordre 2 ou 4. Dans tous les cas, $D_n/ < r^2 >$ est abélien et $D \subset < r^2 >$ (voir section 2.3.4). D'où $D =< r^2 >$.

2.8.3 Une parenthèse : Détermination des groupes d'ordre 8

Soit G un groupe d'ordre 8.

• Si G possède un élément d'ordre 8, alors $G \simeq \mathbb{Z}/8\mathbb{Z}$.

• Si tout élément distinct de 1 est d'ordre 2, la propriété 2.31 montre que G est commutatif et $G \simeq (\mathbb{Z}/2\mathbb{Z})^3$.

• On suppose qu'aucune des conditions précédentes n'est réalisée. Pour x dans $G \setminus \{1\}$, $\circ(x) = 2$ ou 4. On peut choisir $a \neq 1$, $\circ(a) = 4$. Soit $b \notin < a >$.

On a

$$G =< a > \sqcup b < a >= \underbrace{\{1, a, a^2, a^3, b, ab, a^2b, a^3b\}}_{8 éléments\ distincts}$$

donc $G =< a, b >$. On remarque bien que $b^2 \in < a >$, sinon $b^2 \in b < a >$ et $b \in < a >$ (ce qui est absurde). Une première piste pour la construction des groupes d'ordre 8 consiste à examiner successivement les cas :

$$b^2 = 1, \quad b^2 = a, \quad b^2 = a^2, \quad b^2 = a^3.$$

On utilise alors le fait que le sous-groupe $< a >$ d'indice 2 dans G est distingué dans G (cf [4]).

Voici une autre piste : on a clairement

$$ba \in \{ab, a^2b, a^3b\}.$$

- Si $ba = a^2b$, alors $ba^2 = a^2ba = a^2a^2b = b$, ce qui est impossible.

- Si les générateurs a et b de G commutent, par associativité de la loi, G est commutatif. On envisage alors à bon droit le sous-groupe $< a > \times < b >$ de G. Si $\circ(b) \geqslant 3$, alors $< a > \times < b >$ a plus de 8 éléments (ce qui n'est pas), donc $\circ(b) = 2$. On considère $\phi : (\alpha, \beta) \in < a > \times < b > \mapsto \alpha\beta \in G$, qui est un morphisme de groupe (G commutatif), surjectif. Pour des raisons de cardinaux, $G \simeq < a > \times < b > \simeq \mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$.

- Si l'on suppose que $ba = a^3b$ et si l'on tente de remplir la table de G, on est vite coincé parce qu'on ne sait pas ce que fait b^2. Or $\circ(b) = 2$ ou 4.

a) Si $\circ(b) = 2$, alors $b^2 = 1$, $(ab)^2 = abab = aa^3bb = 1$, et on reconnaît le groupe diédral D_4.

b) Si $\circ(b) = 4$, on vérifie (facilement) que $b^2 \notin \{1, a, a^3\}$, donc $b^2 = a^2$, et on construit entièrement la table de G avec les relations $a^4 = 1$, $a^2 = b^2$, et $ba = a^3b$. G est appelé groupe quaternionique $\mathcal{H}_8$.

Pour une réalisation de $\mathcal{H}_8$, on peut considérer les matrices 2×2 complexes I_2, $-I_2$, $a = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$, $-a$, $b = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, $-b$, $c = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$, $-c$ (comme à la page 76).

Remarque : Le 2-groupe D_4 (ou $\mathcal{H}_8$) n'est pas commutatif et son centre Z n'est pas trivial, donc $\circ(Z) = 2$ ou 4. Si $\circ(Z) = 4$, un élément $x \in G \setminus Z$ vérifie : $Z \subsetneq C_x$, donc $C_x = G$, $x \in Z$, ce qui est absurde. Ainsi, $\circ(Z) = 2$, et la section 2.7.3 montre que la probabilité $P(G)$ pour que deux éléments de D_4 (ou $\mathcal{H}_8$) pris au hasard commutent est maximale et égale à :

$$P(D_4) = P(\mathcal{H}_8) = \frac{5}{8}.$$

Exercice 2.22 *Considérons à nouveau l'isomorphisme Φ de groupes de $\mathcal{S}_4$ dans $\mathcal{S}_{\{u,v,w\}} \equiv \mathcal{S}_3$ défini par :*

$$\Phi \ : \ \sigma \mapsto \begin{cases} \{u,v,w\} & \to \{u,v,w\} \\ z & \mapsto \sigma z \sigma^{-1} \end{cases}$$

(où u, v et w ont été définis à la section 2.6.1). Considérons le sous-groupe $U = <(v,w)>$ de $\mathcal{S}_{\{u,v,w\}}$. Le but de l'exercice est de déterminer $\Phi^{-1}(U)$.

a) Montrer que $\Phi^{-1}(U)$ est le centralisateur C_u de u.

b) En déduire que $\Phi^{-1}(U)$ est d'ordre 8.

c) Vérifier que $V_4 \cup \{\tau = (1,2);(3,4)\} \subset \Phi^{-1}(U)$.

d) Montrer que $\Phi^{-1}(U) =< v, \tau > \equiv D_4$.

2.8.4 Cas d'un groupe abélien

On a besoin du lemme suivant :

Lemme 2.13 *Soient H un sous-groupe d'un groupe commutatif G d'ordre n, et $x \in G$ d'ordre k. On pose $L =< H \cup \{x\} >$. Alors il existe un diviseur κ de k tel que : $\circ(L) = \kappa \times \circ(H)$.*

Preuve : Soit $\mathcal{H} = \{m \in \mathbb{Z} \ ; \ x^m \in H\}$. On a $k \in \mathcal{H}$ puisque $x^k = 1$ donc $\mathcal{H}$ est non réduit à $\{0\}$. Par ailleurs, $\mathcal{H}$ est un sous-groupe de $\mathbb{Z}$. Il existe donc $\kappa \in \mathbb{N}$ tel que $\mathcal{H} = \kappa\mathbb{Z}$. On note au passage que κ divise k.

Le sous-groupe L contient tous les éléments de H et toutes les puissances de x, donc $\widetilde{L} = \{hx^m \ ; \ h \in H, m \in \mathbb{Z}\} \subset L$. On vérifie par ailleurs que $\widetilde{L}$ est un sous-groupe de G, qui contient $H \cup \{x\}$, donc $L = \widetilde{L}$.

Pour $m \in \mathbb{Z}$, m s'écrit $m = \kappa q + r$ où $0 \leqslant r < \kappa$. Ainsi

$$u^m = \underbrace{(u^\kappa)^q}_{\in H} u^r,$$

ce qui donne $L = \{hx^m \ ; \ h \in H, 0 \leqslant m < \kappa\}$. On envisage alors :

$$\psi \ : \ (h,m) \in H \times \{0,...,\kappa\} \mapsto hx^m \in L.$$

L'application ψ est clairement surjective. Avec des notations évidentes,

$$hx^m = h'x^{m'} \ \Rightarrow \ x^{m'-m} \in H \ \Rightarrow \ \kappa \mid (m'-m) \ \Rightarrow \ m = m'$$

puisque $m, m' \in \{0,...,\kappa\}$. On vient de prouver que ψ est injective, ce qui donne finalement : $\circ(L) = \kappa \times \circ(H)$. ■

On peut maintenant énoncer :

Propriété 2.33 *Soit G un groupe commutatif d'ordre n et p un diviseur premier de n. Alors G possède un élément d'ordre p.*

Preuve : • *Première méthode* : Soit $x_1, ..., x_n$ les éléments de G. On écrit :

$$G = \left\langle \underbrace{< x_1, ..., x_{n-1} >}_{H} \cup \{x_n\} \right\rangle .$$

D'après le lemme 2.13, n divise $\circ(H) \times \circ(x_n)$. Le raisonnement par descente est clair jusqu'à :

$$n \mid \prod_{x \in G} \circ(x).$$

Soit p un diviseur premier de n. Il existe $x \in G$ tel que $p \mid \circ(x)$. Dans le sous-groupe cyclique $< x >$, on choisit alors à bon droit (propriété 2.6 p. 52) un élément d'ordre p.

• *Seconde méthode (on utilise le théorème d'isomorphisme)* : On appelle x_1, ..., x_n les éléments de G et l_1, ..., l_n leurs ordres. On envisage l'application $f : (y_1, ..., y_n) \mapsto y_1...y_n$ de $< x_1 > \times ... \times < x_n >$ dans (clairement sur) G. Puisque G est commutatif, f est un morphisme de groupes, donc

$$G \equiv (< x_1 > \times ... \times < x_n >)/ker(f).$$

Ainsi $\circ(G) = n \mid l_1...l_n$, $p \mid l_1...l_n$, et il existe un indice i tel que p divise l_i, c'est-à-dire un élément x de G tel que $p \mid \circ(x)$. Dans le sous-groupe cyclique $< x >$, on choisit alors à bon droit un élément d'ordre p (ce qui est possible d'après la propriété 2.6 p. 52). ■

Remarque : Un diviseur premier p de $\circ(G) = n$ est l'ordre d'un certain élément x de G, et l'ordre de cet élément x divise l'exposant m de G (lemme 2.2), donc l'entier premier p est donc aussi un diviseur de m.

Exercice 2.23 *Soient p, q premiers et G un groupe commutatif d'ordre pq. Alors G est cyclique.*

Solution : On choisit a, b dans G tels que $\circ(a) = p$, $\circ(b) = q$. Puisque $ab = ba$ et $p \wedge q = 1$, on a $G =< ab >$.

2.8.5 Groupes d'ordre pair

▶ **Eléments d'ordre 2**

Propriété 2.34 *Soit G un groupe d'ordre $2n$. Alors G possède un élément d'ordre 2.*

Preuve : On définit sur G une relation d'équivalence de la façon suivante :

$$\forall (x,y) \in G \quad x\mathcal{R}y \quad \Leftrightarrow \quad y \in \{x, x^{-1}\}.$$

Soit N_1 le nombre de classes réduites à un point et N_2 le nombre de classes à deux éléments. On a : $2n = N_1 + 2N_2$ donc N_1 est pair. Comme $\bar{1}$ est un singleton, $N_1 \neq 0$, donc $N_1 \geqslant 2$. Il existe donc au moins un élément x de $G \backslash \{1\}$ tel que $x = x^{-1}$. ■

Remarques : α) La relation $\mathcal{R}$ de la preuve ci-dessus permet d'obtenir la congruence de Wilson. Soit p premier. Dans le groupe multiplicatif $(\mathbb{Z}/p\mathbb{Z})^*$, les classes réduites à un singleton sont $\{1\}$ et $\{p-1\}$, et dans le produit $(p-1)!$, on regroupe les facteurs par classe, ce qui donne $(p-1)! \equiv -1 \ (p)$.

β) Et voici une soeur-jumelle de $\mathcal{R}$... Supposons p premier. Si m est un carré dans $(\mathbb{Z}/p\mathbb{Z})^*$, le petit théorème de Fermat donne $m^{\frac{p-1}{2}} = 1$. Si m n'est pas un carré dans $(\mathbb{Z}/p\mathbb{Z})^*$, on considère la relation d'équivalence :

$$x\mathcal{R}_m y \quad \Leftrightarrow \quad y \in \{x, mx^{-1}\}.$$

Les classes suivant $\mathcal{R}_m$ ont toutes deux éléments et en regroupant encore une fois les facteurs de $(p-1)!$ par classe, on obtient :

$$m^{\frac{p-1}{2}} = -1.$$

Propriété 2.35 *Tout groupe fini G vérifiant $P(G) = 5/8$ a un ordre multiple de* 8.

Preuve : [*On utilise ici les résultats et les notations de la section 2.7.*] Si $P(G) = 5/8$, on commence par montrer que tout carré x^2 de G est dans le centre $Z(G)$. En effet, si $x \in Z(G)$, alors $x^2 \in Z(G)$ car $Z(G)$ est un sous-groupe de G. Et si $x \notin Z(G)$, on a

$$\frac{\circ(C_x)}{\circ(Z(G))} = j_x = 2$$

puisque $P(G) = 5/8$. Donc $C_x = Z(G) \sqcup xZ(G)$ et nécessairement $x^2 \notin xZ(G)$ (sinon $x \in Z(G)$).

On a $\circ(G) = m \times \circ(Z(G)) = 4 \times \circ(Z(G))$. Il suffit donc de vérifier que $Z(G)$ possède un élément d'ordre 2. Soit $x \notin Z(G)$. On a $\circ(C_x) = 2 \times \circ(Z(G))$, donc C_x possède un élément u d'ordre 2 (propriété 2.34). Si $u \in Z(G)$, tant mieux. Sinon, on choisit $y \notin C_x$, C_y possède un élément v d'ordre 2. Si $v \in Z(G)$, tant mieux. Dans le cas contraire, on montre que $z = (uv)^2$ est dans le centre $Z(G)$ et d'ordre 2.

- On a $z \neq 1$, sinon $uv = vu$, $C_u = C_v$, $C_x = C_y$, $y \in C_x$, contradiction.

- Tout carré est dans le centre, donc $z \in Z(G)$.

- On a $uz = \underbrace{u^2}_{\in Z(G)} vuv = vuu^2v = vuvu^2 = (vu)^2u = u(vu)^2$ donc $z = (vu)^2$.

- Enfin : $z^2 = (uv)^2\,(vu)^2 = uvuvvuvu = 1$. ■

Pour le plaisir, voici un autre clin d'oeil aux carrés dans un groupe :

Propriété 2.36 *Pour $n \geqslant 3$, $\mathcal{A}_n$ est le seul sous-groupe de $\mathcal{S}_n$ d'indice 2.*

Preuve : Soit H un sous-groupe de $\mathcal{S}_n$ d'indice 2.

- H contient tous les carrés de $\mathcal{S}_n$ (cf. remarque p. 67).

- On sait (cf. lemme 2.5 p. 68) que $\mathcal{A}_n$ est engendré par les 3-cycles. Or, si c est un 3-cycle, $c^3 = 1$ donc $c = \left(c^{-1}\right)^2$, ce qui montre que $\mathcal{A}_n$ est aussi engendré par les carrés de $\mathcal{S}_n$. On conclut en remarquant que le sous groupe H d'indice 2 dans $\mathcal{S}_n$ contient tous les carrés de $\mathcal{S}_n$ donc contient $\mathcal{A}_n$ et pour des raisons de cardinaux, $H = \mathcal{A}_n$. ■

▶ Groupe d'ordre $2p$

Propriété 2.37 *Soit G un groupe d'ordre $2p$ avec p premier. Alors G possède un élément d'ordre p.*

Preuve : • *Première méthode* : Pour $a \in G$, $a \neq 1$, l'ordre de a est $2, p$ ou $2p$. S'il existe $a \in G$ tel que $\circ(a) = 2p$, alors $G \simeq \mathbb{Z}/2p\mathbb{Z}$, et puisque 2 et p sont premiers entre eux, $G \simeq \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. L'élément $(\bar{2}, \bar{0})$, par exemple, est d'ordre p dans $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ donc G possède aussi un élément d'ordre p. Si pour tout $a \in G$, $\circ(a) = 2$, alors $\circ(G)$ est une puissance de 2 d'après la propriété 2.31. Contradiction.

• *Autre méthode possible* : on prend $a \neq 1$, $b \neq 1$, $b \neq a$, et l'on remarque que $< a, b > = \{1, a, b, ab\}$, donc 4 divise $2p$, 2 divise p, ce qui est absurde. ■

▶ Détermination des groupes d'ordre $6 = 2 \times 3$

On suppose G non cyclique. Soit $r \in G$, $r \neq 1$, $\circ(r) = 3$ (il existe d'après la propriété 2.37). On rappelle que le sous-groupe $< r >$ d'indice 2 dans G est distingué (exercice 2.12). Soit $s \in G$, $\circ(s) = 2$ (il existe d'après la propriété 2.34). On montre que $\circ(sr) = 2$. On a $srs = srs^{-1}$ donc $srs \in \{1, r, r^2\}$. Si $srs = 1$, alors $r = 1$! Si $srs = r$, alors $sr = rs$ et l'application

$$\phi\,:\,(\rho, \sigma) \in < r > \times < s > \mapsto \rho\sigma \in G$$

est un morphisme de groupe. Son image contient $\{1, \rho, \rho^2, s\}$ donc son image dont le cardinal divise 6, est nécessairement G. Ainsi ϕ est surjectif, donc $G \simeq \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, donc G est cyclique (puisque $2 \wedge 3 = 1$). Contradiction. Bilan : $srs = r^2$, $srsr = r^3 = 1$. Enfin G contient les 6 éléments distincts $1, s, r, r^2, sr$ et rs, donc $G =< r, s >$. En définitive, $G = D_3$.

Comme D_3 est aussi isomorphe à $\mathcal{S}_3$, on peut énoncer :

Propriété 2.38 *Un groupe d'ordre* 6 *est soit isomorphe à* $\mathbb{Z}/6\mathbb{Z}$ *(et donc commutatif), soit isomorphe à* D_3 *(ou ce qui revient au même à* $\mathcal{S}_3$*, et dans ce cas il n'est pas commutatif).*

Remarque : D_3 (isomorphe à $\mathcal{S}_3$) est le plus petit groupe non commutatif.

Voici une conséquence de la propriété 2.38 :

Propriété 2.39 *Il n'existe aucun sous-groupe d'ordre* 6 *dans* $\mathcal{A}_4$*.*

Preuve : Dans le groupe alterné d'ordre 4, il y a :

a) L'identité Id.

b) Les produits de deux transpositions à supports disjoints qui sont d'ordre 2 et au nombre de 3. On rappelle que ces éléments forment avec Id un sous-groupe de $\mathcal{A}_4$ isomorphe à V_4.

c) Les 3-cycles

$$r_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 4 & 2 \end{pmatrix}, \quad r_2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{pmatrix},$$

$$r_3 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix}, \quad r_4 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix},$$

ainsi que les 3-cycles inverses, tous étant d'ordre 3.

On a listé 12 permutations de $\mathcal{A}_4$, on les a donc toutes.

Si $\mathcal{A}_4$ possède un sous-groupe H d'ordre 6, on a $H \simeq D_3$ puisque H ne possède aucun élément d'ordre 6. On choisit alors un 3-cycle r et un élément s de V_4 tels que : $\circ(rs) = 2$ et $H =< r, s >$. La permutation $\sigma = rs$ appartient donc au sous-groupe V_4, ce qui impose $r = \sigma\ s \in V_4$. Absurde puisque $\circ(r) = 3$. ■

Exercice 2.24

1) Montrer que V_4 *est le seul sous-groupe de* $\mathcal{A}_4$ *d'ordre* 4.

2) Justifier $\mathcal{S}_4/V_4 \equiv \mathcal{S}_3$.

Solution : Soit H un sous-groupe de $\mathcal{A}_4$ d'ordre 4. Tout élément de H est d'ordre 1, 2 ou 4, et puisque $\mathcal{A}_4$ ne possède aucun élément d'ordre 4, tout élément de H est involutif. Par nécessité, $H = V_4$.
Par la formule des indices, $\mathcal{S}_4/V_4$ a 6 éléments, donc est isomorphe à $\mathbb{Z}/6\mathbb{Z}$ ou $\mathcal{S}_3$. On pose $x = (1,2)$ et $y = (2,3)$. Le commutateur

$$xyx^{-1}y^{-1} = (1,2,3)(1,2,3) = (1,3,2)$$

est un 3-cycle donc $xy(yx)^{-1} \notin V_4$, $\overline{xy} \neq \overline{yx}$. Ainsi, $\mathcal{S}_4/V_4$ n'est pas commutatif, d'où : $\mathcal{S}_4/V_4 \equiv \mathcal{S}_3$.

Exercice 2.25 *Montrer que le groupe $Aut(\mathcal{S}_3)$ des automorphismes de $\mathcal{S}_3$ est isomorphe à $\mathcal{S}_3$.*

Solution : Soit $\mathcal{T} = \{(1,2);(1,3);(2,3)\}$ l'ensemble des transpositions de $\mathcal{S}_3$. Puisque les automorphismes de groupes conservent la période (ou l'ordre) des éléments du groupe et puisque $\mathcal{T}$ est exactement les permutations d'ordre 2 de $\mathcal{S}_3$, on envisage à bon droit le morphisme[2]

$$\Phi : \ \phi \in Aut(\mathcal{S}_3) \mapsto [\tau \in \mathcal{T} \mapsto \phi(\tau)]$$

de $Aut(\mathcal{S}_3)$ dans le groupe (isomorphe à $\mathcal{S}_3$) des permutations de $\mathcal{T}$.
Si $\phi \in ker(\Phi)$, alors la restriction de ϕ à $\mathcal{T}$ est l'identité, et puisque $\mathcal{T}$ engendre $\mathcal{S}_3$, ϕ est l'identité. Ainsi Φ est injective, donc $Aut(\mathcal{S}_3)$ est isomorphe au sous-groupe $\Phi(Aut(\mathcal{S}_3))$ de $\mathcal{S}_3$. A présent, le sous-groupe $Int(\mathcal{S}_3)$ de $Aut(\mathcal{S}_3)$ des automorphismes intérieurs de $\mathcal{S}_3$ est isomorphe à $\mathcal{S}_3/Z(\mathcal{S}_3) \equiv \mathcal{S}_3$ puisque le centre $Z(\mathcal{S}_3)$ est trivial (propriété 2.17). Il vient $\sharp(Aut(\mathcal{S}_3)) \geqslant \sharp(Int(\mathcal{S}_3)) = 6$. Avec ce qui précède, $Aut(\mathcal{S}_3) = Int(\mathcal{S}_3) \equiv \Phi(Aut(\mathcal{S}_3)) = \mathcal{S}_3$.

▶ Détermination des groupes d'ordre $2p$ (où p premier impair)

On choisit dans G un élément r d'ordre p, et un élément s d'ordre 2. On a $s \notin < r >$ car 2 ne divise pas p. Avec $< r >$ d'indice 2 dans G, on peut écrire : $G/< r > = \{< r >, s < r >\}$, $G = \{1, r, .., r^{p-1}\} \cup \{s, sr, .., sr^{p-1}\}$, et $G = < r, s >$.

- Si G est cyclique, $G \simeq \mathbb{Z}/2p\mathbb{Z} \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$.

- On suppose G non cyclique. Pour $\sigma \notin < r >$, σ s'écrit $\sigma = sr^k$ (avec $0 \leqslant k \leqslant p-1$). Si $\sigma^2 \neq 1$, alors $\sigma^2 = (sr^k s)\, r^k = (sr^k s^{-1})\, r^k$ et avec $< r >$ distingué, on a $\sigma^2 = r^i$ avec $1 \leqslant i \leqslant p-1$. Avec $p = 2q+1$, on écrit : $\sigma^{2q} = \sigma^{p-1} = r^{iq}$ donc $\sigma r^{iq} = 1$, $\sigma \in < r >$. Contradiction.

[2]Action naturelle de $Aut(\mathcal{S}_3)$ sur $\mathcal{T} = \{(1,2);(1,3);(2,3)\}$.

Bilan : $\forall \sigma \notin < r > \quad \sigma^2 = 1$. En particulier, $(sr)^2 = 1$ et on reconnaît le groupe diédral d'indice p.

On a montré :

Propriété 2.40 *Un groupe d'ordre $2p$ (où p est un nombre premier impair) est soit isomorphe à $\mathbb{Z}/2p\mathbb{Z}$ (donc aussi à $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, et donc commutatif), soit isomorphe à D_p (et dans ce cas il n'est pas commutatif).*

Remarque : Le groupe D_p est un groupe non cyclique dont tous les sous-groupes propres sont cycliques.

▶ **Sous-groupe d'indice 2 et unicité**

Propriété 2.41 *Soit G un groupe fini et H un sous-groupe distingué de G d'ordre m. On suppose que : $m \wedge [G : H] = 1$. Alors H est l'unique sous-groupe (distingué) de G d'ordre m.*

Preuve : On note Π le morphisme surjectif $x \mapsto \overline{x}$ de G sur G/H. Si K désigne un sous-groupe de G d'ordre m, on envisage la restriction de Π à K, qu'on continue à noter Π. Puisque $\Pi(K)$ est un sous-groupe de G/H, $\circ(\Pi(K))$ divise $[G : H]$. Par ailleurs, puisque $K/ker(\Pi) \equiv \Pi(K)$, $\circ(\Pi(K))$ divise $\circ(K) = m$. Or $m \wedge [G : H] = 1$, donc $\circ(\Pi(K)) = 1$, $\Pi(K) = \{H\}$, $K \subset H$, et finalement $K = H$. ■

Remarque : Peut-on se passer de l'hypothèse "H distingué" dans la propriété 2.41 ? Non, dans $\mathcal{S}_3$ de cardinal 2×3, considérer les sous-groupes $< (1, 2) >$ et $< (2, 3) >$ qui sont d'ordre 2.

Conséquences : α) On retrouve le fait que V_4 distingué dans $\mathcal{A}_4$ est l'unique sous-groupe de $\mathcal{A}_4$ d'ordre 4.

β) Si m est un entier naturel impair, alors le sous-groupe $< r >$ des rotations du groupe diédral $D_m =< r, s >$ d'indice m est l'unique sous-groupe de D_m d'indice 2.

2.8.6 Cas général

Propriété 2.42 ***(Théorème de Cauchy)*** *Soit G un groupe fini d'ordre n et p un diviseur premier de n. Alors G possède un élément d'ordre p.*

Preuve : [*cf.* [4]] On raisonne par récurrence sur l'ordre n de G. Si $n = 2$, c'est évident. On suppose la propriété vraie pour tout groupe d'ordre $m < n$. Soit p un diviseur premier de n. Si G est commutatif, c'est acquis (propriété 2.33). On suppose donc désormais : $Z(G) \neq G$.

- S'il existe $x \in G \setminus Z(G)$ tel que $p \mid \circ(C_x)$, on applique à bon droit l'hypothèse de récurrence à C_x ($x \notin Z(G) \Rightarrow C_x \subsetneq G$) : il existe dans C_x (et donc dans G) un élément d'ordre p.

- Si pour tout $x \in G \setminus Z(G)$ p ne divise pas $\circ(C_x)$, p (premier) qui divise $n = \circ(C_x) \times [G : C_x]$, divise $[G : C_x]$ et d'après l'équation aux classes, p divise $\circ(Z(G))$. D'après l'hypothèse de récurrence, $Z(G)$ donc G possède alors un élément d'ordre p. ■

On peut montrer mieux :

Propriété 2.43 *Soit G un groupe fini de cardinal n et p un nombre premier. Si $p^m \mid n$ avec $m \in \mathbb{N}$, alors G contient un sous-groupe d'ordre p^m.*

Application 2.2 *Soit G un groupe fini et $p > 0$ un nombre premier. Les deux propriétés suivantes sont équivalentes :*
(1) $\circ(G)$ est une puissance de p (i.e. G est un p-groupe).
(2) Tout élément de G a pour ordre une puissance de p.

Preuve : (1) $\Rightarrow$ (2) est une conséquence immédiate du théorème de Lagrange. On suppose maintenant que l'ordre de tout élément de G est une puissance de p. Soit q un diviseur premier de $\circ(G)$. Avec Cauchy, G possède un élément d'ordre q, donc q s'écrit p^α, et puisque q est premier, $q = p$. D'où le résultat. ■

Application 2.3 *Il n'y a qu'un seul sous-groupe d'ordre* 15*, à savoir* $\mathbb{Z}/15\mathbb{Z}$*.*

Preuve : Avec Cauchy, on choisit un élément a d'ordre 5, et un élément b d'ordre 3. Puisque 3 est le plus petit diviseur premier de l'ordre de G, le sous-groupe $< a >$ est distingué dans G (propriété 2.14 p. 63).
Ainsi, $b^{-1}ab = b^2ab = a$, ou $b^2ab = a^2$, ou $b^2ab = a^3$, ou $b^2ab = a^4$.

- Si $b^2ab = a$, alors $ab = ba$ et puisque $\circ(a) \wedge \circ(b) = 1$, ab est d'ordre $5 \times 3 = 15$: G est isomorphe à $\mathbb{Z}/15\mathbb{Z}$.

- Si $b^2ab = a^2$, alors $\left(b^2ab\right)^3 = a$, encore $b^2a^3b = a$. Par ailleurs, $b^2ab = a^2$ donne $ba^2b^2 = a$. Ainsi, $b^2a^3b = ba^2b^2$, $a^3 = b^2a^2b$, puis $(a^3)^2 = b^2a^4b$, c-à-d $a = b^2a^4b$. Il vient $b^2a^4b = b^2a^3b$, $a^4 = a^3$, $a = 1$, ce qui est faux.

- Si $b^2ab = a^3$, alors $\left(b^2ab\right)^2 = a^6 = a$, encore $b^2a^2b = a$. Par ailleurs, $b^2ab = a^3$ donne $ba^3b^2 = a$. Ainsi, $b^2a^2b = ba^3b^2$, $b^4a^2b = a^3b^2$, $ba^2b = a^3b^2$, $ba^2b^2 = a^3$, puis $(a^3)^2 = a = ba^4b^2$. Il vient $ba^4b^2 = ba^3b^2$, $a^4 = a^3$, $a = 1$, ce qui est faux.

- Si $b^2ab = a^4$, alors $(b^2ab)^4 = b^2a^4b = a$. Par ailleurs, $b^2ab = a^4$ donne $a = ba^4b^2$. Ainsi, $b^2a^4b = ba^4b^2$, $a^4b = b^2a^4b^2$, $a^4b^2 = b^2a^4$. Or a^4 (comme a) est d'ordre 5 et b^2 (comme b) est d'ordre 3. Il vient $\circ(a^4b^2) = 15$, puis G

cyclique. Le groupe G est donc commutatif, donc $b^2ab = a$, $a^4 = a$, ce qui est faux. ■

Application 2.4 *(avec le concours de Michel Vieumelen)*
Soit $p > 2$ un nombre premier et G un groupe d'ordre $p + 1$. On note $Aut(G)$ le groupe des automorphismes de G. Le nombre p divise l'ordre de $Aut(G)$ si, et seulement si, il existe $n > 1$ tel que $G \simeq (\mathbb{Z}/2\mathbb{Z})^n$.

Preuve : • Si p premier divise l'ordre de $Aut(G)$, d'après Cauchy, le groupe $Aut(G)$ possède un élément φ d'ordre p. On pose $\Gamma =< \varphi >$.
Sur $X = G \setminus \{1\}$, on considère alors la relation d'équivalence [3] :

$$x\mathcal{T}y \quad \Leftrightarrow \quad \exists 0 \leqslant k \leqslant p-1 \quad \varphi^k(x) = y.$$

Pour $x \in X$, on pose $\Gamma_x = \{\psi \in \Gamma \ ; \ \psi(x) = x\}$. On vérifie que Γ_x est un sous-groupe de Γ, donc $\Gamma_x = \{Id\}$, ou $\Gamma_x = \Gamma$. Si pour tout $x \in X$, $\Gamma_x = \Gamma$, alors

$$\forall x \in X \quad \forall \psi \in \Gamma \quad \psi(x) = x,$$

donc $\varphi = Id$. Contradiction. Ainsi

$$\exists a \in X \quad \Gamma_a = \{Id\}.$$

Les p éléments $a, \varphi(a)$, ..., $\varphi^{p-1}(a)$ sont alors distincts dans X. G étant d'ordre pair, X possède un élément $b = \varphi^l(a)$ d'ordre 2. Pour x quelconque dans X, x s'écrit $x = \varphi^k(a) = \varphi^{k-l}(b)$. On a alors $x^2 = \varphi^{k-l}(b^2) = 1$, ce qui assure que tout élément de G est involutif et prouve que $G \simeq (\mathbb{Z}/2\mathbb{Z})^n$ (on pourra reprendre la preuve ci-dessus en utilisant le langage éclairant des actions de groupes).

• Pour la réciproque, on remarque que $Aut((\mathbb{Z}/2\mathbb{Z})^n) = GL_n(\mathbb{Z}/2\mathbb{Z})$ possède $(2^n - 1)(2^n - 2)...(2^n - 2^{n-1})$ éléments (lemme 2.11 p. 82), donc on a bien

$$p = 2^n - 1 \mid \circ\left(Aut((\mathbb{Z}/2\mathbb{Z})^n)\right). \ ■$$

Références

[1] M. Alessandri, Thèmes de géométrie, chapitre 1, Dunod.

[2] RMS mai-juin 1996, Examens oraux, Vuibert.

[3] M. Quercia, F. Ranty, Problèmes corrigés de mathématiques supérieures, Ellipses.

[4] A. Bouvier, D. Richard, Groupes, Hermann.

[5] P. Meunier, Exercices d'algèbre et d'analyse, Tome 2, PUF.

[6] D. Guin, Algèbre, Tome 1, Belin et Editions espaces 34.

[3] Action naturelle de $< \varphi >$ sur X.

Chapitre 3

Introduction aux espaces projectifs

Introduction aux espaces projectifs, preuves des théorèmes de Pappus et de Desargues, dualité.

(Dany-Jack Mercier[1])

Résumé : Voici une introduction aux espaces projectifs. Après avoir rendu compte du lien intime qui existe entre les espaces affines et les espaces projectifs, on définit une topologie sur le projectif, puis on justifie l'emploi de ces espaces en montrant que ce cadre de travail simplifie les démonstrations des théorèmes de Pappus et de Desargues en nous évitant d'avoir à envisager les nombreux cas particuliers qui apparaissent très vite en géométrie affine. On achève cette incursion projective en montrant comment la dualité permet de déduire mécaniquement des énoncés de théorèmes, et en nous intéressant aux homographies. Le lecteur qui désire atteindre rapidement les preuves des théorèmes de Pappus et Desargues peut sauter la Section 3.6.
Cet article convient parfaitement pour découvrir les espaces projectifs. Il devrait aussi permettre aux candidats aux CAPES et aux agrégations de prendre du recul en leur donnant l'occasion de :
- mettre en oeuvre des résultats vectoriels classiques,
- compléter une culture mathématique générale,
- mieux comprendre deux jolis théorèmes dans des environnements géométriques différents.

[1] CRREF, IREM de Guadeloupe, dany-jack.mercier@univ-ag.fr.

3.1 Introduction

▶ Dans un plan affine, deux droites peuvent de pas se couper. Cela nous oblige à envisager de nombreux cas particuliers qui alourdissent les démonstrations.

▶ Dans un espace affine de dimension finie sur un corps commutatif K, la dimension de l'espace affine $\text{Aff}\,(F \cup G)$ engendré par la réunion de deux sous-espaces affines est facile à calculer, mais nécessite d'envisager deux cas. Notons $F = A + \overrightarrow{F}$ et $G = B + \overrightarrow{G}$, et utilisons le Th. 7 de [1].

- Si $\overrightarrow{AB} \in \overrightarrow{F} + \overrightarrow{G}$, alors $F \cap G \neq \varnothing$, $F \cap G$ est un sous-espace affine de direction $\overrightarrow{F} \cap \overrightarrow{G}$, et $\text{Aff}\,(F \cup G)$ est de direction $\text{Vect}(\overrightarrow{F} \cup \overrightarrow{G}) = \overrightarrow{F} + \overrightarrow{G}$. On obtient

$$\begin{aligned}
\dim(\text{Aff}\,(F \cup G)) &= \dim(\overrightarrow{F} + \overrightarrow{G}) \\
&= \dim \overrightarrow{F} + \dim \overrightarrow{G} - \dim(\overrightarrow{F} \cap \overrightarrow{G}) \\
&= \dim F + \dim G - \dim(F \cap G).
\end{aligned}$$

- Si $\overrightarrow{AB} \notin \overrightarrow{F} + \overrightarrow{G}$, alors $F \cap G = \varnothing$ et $\text{Aff}\,(F \cup G)$ est de direction $\text{Vect}(\overrightarrow{F} \cup \overrightarrow{G} \cup \{\overrightarrow{AB}\})$, donc

$$\begin{aligned}
\dim(\text{Aff}\,(F \cup G)) &= \dim(\overrightarrow{F} + \overrightarrow{G} + K\overrightarrow{AB}) \\
&= \dim(\overrightarrow{F} + \overrightarrow{G}) + 1 \\
&= \dim \overrightarrow{F} + \dim \overrightarrow{G} - \dim(\overrightarrow{F} \cap \overrightarrow{G}) + 1 \\
&= \dim F + \dim G - \dim(\overrightarrow{F} \cap \overrightarrow{G}) + 1.
\end{aligned}$$

La formule $\dim(\text{Aff}\,(F \cup G)) = \dim F + \dim G - \dim(F \cap G)$ n'est pas toujours vraie. Peut-on imaginer un espace ressemblant fondamentalement à un espace affine, mais où deux sous-espaces s'interceptent toujours ? (réponse au Théorème 3.3)

▶ La FIG. 3.1 représente deux droites sécantes D et D', ainsi qu'un point A n'appartenant ni à D, ni à D'.
La fonction f qui à tout point M de D associe l'éventuel point N de D' tel que M, A, N soient alignés, n'est pas définie sur tout D. Le point M_0, intersection de D et de la parallèle à D' passant par A, n'a pas d'image, tandis que le point N_0, intersection de D' et de la parallèle à D passant par A, n'a pas d'antécédent.
A cause des points M_0 et N_0, la fonction $f : D \to D'$ n'est ni une application, ni une surjection.

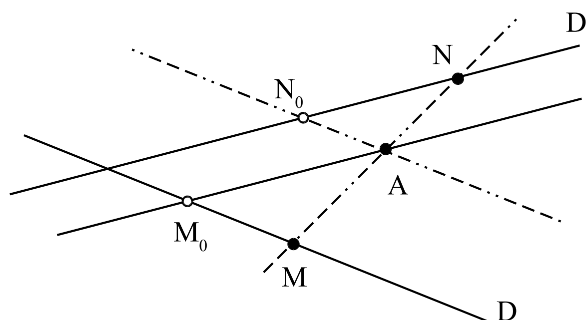


FIG. 3.1 – $f : M \longmapsto N$

Pour traiter cette pathologie, l'idée consiste à adjoindre un nouveau point à chacune des droites D et D'. Ces points, dits "points à l'infini" et notés ∞_D et $\infty_{D'}$, appartiendront à n'importe quelle droite ayant la même direction que D ou D'. Il suffit alors de poser $f(\infty_D) = N_0$ et $f(M_0) = \infty_{D'}$ pour que l'application $f : D \cup \{\infty_D\} \to D' \cup \{\infty_{D'}\}$ soit bien définie et bijective. Une "droite" ne peut-elle pas être considérée comme une droite affine standard D augmentée d'un point à l'infini qui caractérise sa direction (le fameux point ∞_D) ?

► C'est l'ingénieur militaire français Girard Desargues[2] (1591-1661) qui le premier définit le plan projectif comme un plan "usuel" auquel on rajoute des points à l'infini correspondant, chaque fois, à des directions de droites, deux droites parallèles se coupant alors toujours en un point à l'infini.
Si D est une droite du plan usuel E, et si l'on note ∞_D sa "direction", le plan projectif $\mathbb{P}(E)$ est l'ensemble

$$\mathbb{P}(E) = E \cup D_\infty$$

où $D_\infty = \{\infty_D \,/\, D \text{ droite de } E\}$ représente la "droite à l'infini" (formée de tous les points "à l'infini").

La FIG. 3.2 montre comment on peut imaginer un plan projectif : l'intérieur (strict) du disque que l'on a dessiné représente le plan affine usuel, et ce disque est donc, en fait, de diamètre "infini". Des droites de même direction

[2]Cet ingénieur continua les travaux des grecs Apollonius de Perge (3ème s. av. J.-C.), Ménélaüs d'Alexandrie (1er s. ap. J.-C.) et Pappus d'Alexandrie (4ème s. ap. J.-C.) sur les sections coniques (intersections d'un cône et d'un plan) et les problèmes de perspective (par projection centrale). Les théorèmes affines qui portent son nom s'énoncent d'une seule manière si l'on se place dans le plan projectif.

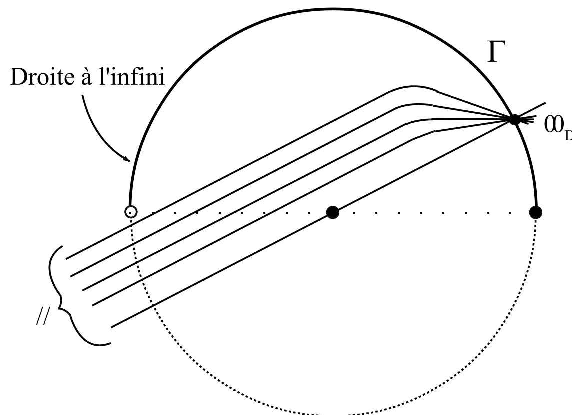


FIG. 3.2 – Une représentation du plan projectif

définissent un unique point ∞_D situé sur le demi-cercle supérieur Γ (ouvert à gauche, fermé à droite) dessiné sur la frontière du disque. Ce point, qui caractérise toutes les droites possédant une direction commune, appartient à chacune d'elles. Maintenant, deux droites parallèles et distinctes D et D' se couperont toujours en un point ∞_D que l'on dit être "à l'infini". Le demi-cercle Γ formé de tous les points à l'infini est naturellement appelé **droite à l'infini**.
Dans la Section suivante, nous allons proposer une définition plus générale et plus rigoureuse d'un espace projectif, mais la visualisation précédente doit accompagner notre incursion dans le domaine projectif. C'est elle qui nous permet "d'imaginer" notre plan affine comme "baignant" dans un plan projectif obtenu en rajoutant une nouvelle droite (la droite à l'infini).

3.2 Définitions

Dans toute la suite, K désigne un corps commutatif.

Définition 3.1 *Soit E un espace vectoriel de dimension fini sur K.* ***L'espace projectif*** *associé à E est l'ensemble quotient de $E\backslash\{0\}$ par la relation d'équivalence*

$$x\mathcal{R}y \;\Leftrightarrow\; \exists\lambda\in K^* \quad x=\lambda y.$$

On le note $\mathbb{P}(E)$, de sorte que $\mathbb{P}(E)=(E\backslash\{0\})/\mathcal{R}$.

La classe d'équivalence $\dot{x}$ de $x\in E\backslash\{0\}$ suivant $\mathcal{R}$ est la droite $\mathrm{Vect}\,(x)$ engendrée par x privée du vecteur nul. La surjection canonique $p:E\backslash\{0\}\to\mathbb{P}(E)$ est l'application qui à un vecteur non nul x de E associe sa classe $\dot{x}$ suivant

la relation $\mathcal{R}$. On confondra pratiquement cette surjection canonique avec la fonction $p : E \to \mathbb{P}(E)$ qui reste surjective, tout en n'étant plus définie sur E tout entier.

Comme nous l'avons noté, chaque classe $\dot{x}$ de $\mathbb{P}(E)$ représente une droite de E privée du vecteur nul. L'espace projectif $\mathbb{P}(E)$ est donc en bijection naturelle avec *l'ensemble des droites vectorielles de* E, autrement dit avec la grassmannienne $G_{E,1}$ de dimension 1 de E (on appelle **grassmannienne de dimension** m de E, et l'on note $G_{E,m}$, l'ensemble des sous-espaces vectoriels de dimension m de E).
On peut écrire
$$\mathbb{P}(E) = G_{E,1}.$$

Définition 3.2 *La* ***dimension*** *de* $\mathbb{P}(E)$ *est* $\dim \mathbb{P}(E) = \dim E - 1$. *On dit que* $\mathbb{P}(E)$ *est une droite (resp. un plan) projectif si* $\dim \mathbb{P}(E) = 1$ *(resp.* 2*). Si* $n \in \mathbb{N}^*$, *l'ensemble* $\mathbb{P}(K^{n+1})$ *est appelé* ***espace projectif standard de dimension*** n *sur* K. *On le note indifféremment* $\mathbb{P}^n = \mathbb{P}^n(K) = \mathbb{P}(K^{n+1})$ *si aucune confusion est possible. Un espace projectif est dit* ***réel*** *si* $K = \mathbb{R}$, ***complexe*** *si* $K = \mathbb{C}$. *Si* $\overrightarrow{F}$ *est un sous-espace vectoriel de* E, *l'ensemble* $F = p(\overrightarrow{F})$ *est appelé* ***sous-espace projectif*** *de* $\mathbb{P}(E)$ *associé à* $\overrightarrow{F}$. *On pose alors* $\dim F = \dim \overrightarrow{F} - 1$.

▶ $p(\overrightarrow{F}) = \mathbb{P}(\overrightarrow{F})$ est bien l'espace projectif associé à $\overrightarrow{F}$.

▶ Poser $\dim F = \dim \overrightarrow{F} - 1$ est agréable, car correspond à notre intuition. En effet, si $\overrightarrow{F}$ est une droite vectorielle, $F = p(\overrightarrow{F})$ est un singleton auquel on attribue la dimension $\dim \overrightarrow{F} - 1 = 0$, ce qui nous convient parfaitement !

Convention : L'ensemble vide peut être considéré comme un sous-espace projectif de $\mathbb{P}(E)$ associé au sous-espace vectoriel $\{0\}$. Après tout, p est une fonction de E sur $\mathbb{P}(E)$ (définie sur $E \setminus \{0\}$) et $p(\{0\}) = \varnothing$. La dimension de l'ensemble vide est alors $\dim \varnothing = \dim \{0\} - 1 = -1$.

3.3 Propriétés

Théorème 3.1 *Une intersection* $\bigcap_{i\in I} F_i$ *de sous-espaces projectifs* F_i *associés à des sous-espaces vectoriels* $\overrightarrow{F_i}$ *est un sous-espace projectif associé à* $\bigcap_{i\in I} \overrightarrow{F_i}$.

Preuve : Si $F_i = p(\overrightarrow{F_i})$ pour tout $i \in I$, où les $\overrightarrow{F_i}$ représentent des sous-espaces vectoriels de E, alors $\bigcap_{i\in I} F_i = \bigcap_{i\in I} p(\overrightarrow{F_i}) = p(\bigcap_{i\in I} \overrightarrow{F_i})$. En effet, si l'inclusion

$\bigcap_{i\in I} p(\overrightarrow{F_i}) \supset p(\bigcap_{i\in I} \overrightarrow{F_i})$ est toujours vraie, on remarque que, réciproquement, si $M \in \bigcap_{i\in I} p(\overrightarrow{F_i})$, alors

$$\forall i \in I \quad \exists x_i \in \overrightarrow{F_i} \setminus \{0\} \quad M = p(x_i),$$

de sorte que pour n'importe quels indices i et j, $M = p(x_i) = p(x_j)$. Ainsi

$$\forall i \neq j \quad \exists \lambda \in K^* \quad x_j = \lambda x_i.$$

Si l'on fixe l'indice j, on peut maintenant affirmer que $x_j \in \bigcap_{i\in I} \overrightarrow{F_i}$, et donc que $M = p(x_j) \in p(\bigcap_{i\in I} \overrightarrow{F_i})$. ■

La stabilité des sous-espaces projectifs par intersection permet de définir le **sous-espace projectif engendré par une partie** A de $\mathbb{P}(E)$. C'est, par définition, l'intersection

$$\bigcap_{\substack{F \text{ s.e.p.} \\ F\supset A}} F$$

de tous les sous-espaces projectifs contenant A, ensemble que l'on note $\mathrm{Proj}(A)$. Bien entendu, $\mathrm{Proj}(A)$ est le plus petit sous-espace projectif de $\mathbb{P}(E)$ contenant A (pour la relation d'ordre $\subset$ dans $\mathbb{P}(E)$).

Théorème 3.2 *Si $A \subset \mathbb{P}(E)$, le sous-espace projectif engendré par A est associé au sous-espace vectoriel engendré par $p^{-1}(A)$. Autrement dit*

$$\mathrm{Proj}(A) = p\left(\mathrm{Vect}\left(p^{-1}(A)\right)\right).$$

Preuve : Si $F = p(\overrightarrow{F})$ est un sous-espace projectif qui contient A, on a $\overrightarrow{F} = p^{-1}(F) \cup \{0\} \supset p^{-1}(A)$. Réciproquement, tout sous-espace vectoriel $\overrightarrow{F}$ de E tel que $\overrightarrow{F} \supset p^{-1}(A)$ définit un sous-espace projectif $P = p(\overrightarrow{F})$ tel que $F \supset A$. Ainsi

$$\mathrm{Proj}(A) = \bigcap_{\substack{F \text{ s.e.p.} \\ F\supset A}} F = \bigcap_{\substack{\overrightarrow{F} \text{ s.e.v.} \\ \overrightarrow{F}\supset p^{-1}(A)}} p(\overrightarrow{F}) = p\Big(\bigcap_{\substack{\overrightarrow{F} \text{ s.e.v.} \\ \overrightarrow{F}\supset p^{-1}(A)}} \overrightarrow{F}\Big) = p\left(\mathrm{Vect}\left(p^{-1}(A)\right)\right). \blacksquare$$

Théorème 3.3 ***(Théorème des dimensions)***

1) Si F et G sont deux sous-espaces projectifs de $\mathbb{P}(E)$,

$$\dim(\mathrm{Proj}(F \cup G)) = \dim F + \dim G - \dim(F \cap G).$$

2) En particulier, si $\dim F + \dim G \geq \dim \mathbb{P}(E)$, l'intersection $F \cap G$ n'est pas vide, et deux droites du plan projectif sont toujours sécantes.

Preuve : 1) Si F et G sont associés aux sous-espaces vectoriels $\overrightarrow{F}$ et $\overrightarrow{G}$,

$$\begin{aligned}\dim(\mathrm{Proj}(F \cup G)) &= \dim(\overrightarrow{F} + \overrightarrow{G}) - 1 \\ &= \dim \overrightarrow{F} + \dim \overrightarrow{G} - \dim(\overrightarrow{F} \cap \overrightarrow{G}) - 1 \\ &= \dim F + \dim G - \dim(F \cap G).\end{aligned}$$

2) Si $\dim F + \dim G \geq \dim \mathbb{P}(E)$, la formule entraîne $\dim(F \cap G) \geq 0$ d'où $F \cap G \neq \varnothing$. Enfin deux droites D et D' du plan projectif sont de dimension 1, donc vérifient $\dim D + \dim D' \geq 2$. L'intersection $D \cap D'$ n'est donc jamais vide : ce sera le sous-espace projectif associé à $\overrightarrow{D} \cap \overrightarrow{D}'$. ■

3.4 Coordonnées homogènes

3.4.1 Définitions

Définition 3.3 *Soit $e = (e_0, ..., e_n)$ une base de l'espace vectoriel E. Les* ***coordonnées homogènes*** *d'un point M de $\mathbb{P}(E)$ relativement à cette base sont, par définition, les coordonnées d'un des vecteurs x tel que $p(x) = M$ dans la base e. De telles coordonnées sont définies à un coefficient multiplicatif non nul près. On note $(x_0 : x_1 : ... : x_n)$ un tel système de coordonnées homogènes de M. Ainsi*

$$M = (x_0 : x_1 : ... : x_n) \Leftrightarrow M = p(\sum_{i=0}^{n} x_i e_i).$$

Définition 3.4 *Les points P_0, P_1, ..., P_k de $\mathbb{P}(E)$ sont* ***projectivement indépendants*** *s'ils proviennent de vecteurs linéairement indépendants. On dit alors aussi que la famille $(P_0, P_1, ..., P_k)$ est* ***projectivement libre****.*

Définition 3.5 *On suppose $\dim \mathbb{P}(E) = n$. Un* ***repère projectif*** *de $\mathbb{P}(E)$ est la donnée d'une $(n+2)$-liste $(P_0, P_1, ..., P_{n+1})$ de points de $\mathbb{P}(E)$ tels qu'il existe une base $e = (e_0, ..., e_n)$ de E avec*
1) $\forall i \in \{0, 1, ..., n\}$ $P_i = p(e_i) = (0 : ... : 0 : 1 : 0 : ... : 0)$ (le 1 est à la i-ième place),
2) $P_{n+1} = (1 : 1 : ... : 1)$.
Les $P_0, P_1, ..., P_n$ sont les ***points base****, et P_{n+1} est le* ***point unité****.*

La donnée d'un repère projectif $(P_0, P_1, ..., P_{n+1})$ de $\mathbb{P}(E)$ permet de reconstituer très précisément une base e de E et d'introduire les coordonnées homogènes d'un point quelconque dans cette base (que l'on appellera coordonnées homogènes dans le repère projectif). En effet, si $e = (e_0, ..., e_n)$ et

$e' = (e'_0, ..., e'_n)$ sont deux bases de E vérifiant les conditions de la définition précédente, pour tout $i \in \{0, 1, ..., n\}$,

$$P_i = p(e_i) = p(e'_i) \;\Leftrightarrow\; \exists \lambda_i \in K^* \quad e'_i = \lambda_i e_i,$$

et

$$P_{n+1} = p(\sum_{i=0}^{n} e_i) = p(\sum_{i=0}^{n} e'_i) \quad \Leftrightarrow \quad \exists \lambda \in K^* \quad \sum_{i=0}^{n} e'_i = \lambda \sum_{i=0}^{n} e_i$$

par conséquent $\lambda_i = \lambda$ pour tous les $i \in \{0, 1, ..., n\}$. Les bases e' et e sont proportionnelles, et l'on peut indifféremment choisir l'une ou l'autre pour déterminer les coordonnées homogènes d'un point du projectif : en effet, si M admet les coordonnées $(x_0 : x_1 : ... : x_n)$ dans e, il admettra les coordonnées $(x_0/\lambda : x_1/\lambda : ... : x_n/\lambda)$ dans e', et l'on sait que des coordonnées homogènes proportionnelles définissent le même point projectif.

Théorème 3.4 *Une famille* $(P_0, P_1, ..., P_{n+1})$ *de* $n+2$ *points de l'espace projectif* $\mathbb{P}(E)$ *est un repère projectif si et seulement si toute sous-famille de* $n+1$ *points est projectivement libre (c'est-à-dire si aucun hyperplan ne contient* $n+1$ *points parmi les* $P_0, P_1, ..., P_{n+1}$*).*

Preuve : La condition est nécessaire. Réciproquement, si les points $P_i = p(e_i)$ sont donnés tels que $n+1$ vecteurs quelconques de la famille $(e_0, ..., e_{n+1})$ sont linéairement indépendants, il s'agit de déterminer une base $e' = (e'_0, ..., e'_n)$ de E telle que

$$\begin{cases} \forall i \in \{0, 1, ..., n\} \quad P_i = p(e_i) = p(e'_i) \\ P_{n+1} = p(e_{n+1}) = p(e'_0 + ... + e'_n). \end{cases}$$

Autrement dit, il faut trouver des scalaires λ et λ_i tels que $e'_i = \lambda_i e_i$ si $0 \leq i \leq n$ et $e_{n+1} = \lambda(e'_0 + ... + e'_n)$.
Le vecteur e_{n+1} s'exprime dans la base $(e_0, ..., e_n)$, disons par $e_{n+1} = \sum_{i=0}^{n} \xi_i e_i$, et la condition à réaliser est

$$\sum_{i=0}^{n} \xi_i e_i = \lambda(\lambda_0 e_0 + ... + \lambda_n e_n)$$

ou encore

$$\forall i \in \{0, 1, ..., n\} \quad \lambda\lambda_i = \xi_i.$$

Il suffit de prendre $\lambda = 1$ et $\lambda_i = \xi_i$ pour tout $0 \leq i \leq n$. C'est possible puisqu'aucun des coefficients ξ_i est nul (en effet, si $\xi_{i_0} = 0$ alors le système $(e_0, e_1, ..., \widehat{e_{i_0}}, ..., e_{n+1})$ est lié, ce qui est absurde). ■

Ainsi **:**

- Trois points A, B, C de la droite projective forment un repère projectif si, et seulement si, ils sont distincts deux à deux.

- Quatre points A, B, C, D du plan projectif forment un repère projectif si et seulement si trois quelconques d'entre eux ne sont jamais alignés.

Dans le plan projectif, une démonstration commence souvent par le choix d'un repère projectif "intéressant", bien adapté à la figure sur laquelle on travaille. Cela permet d'utiliser des équations de droites simples dans ce repère. Cette technique sera mise en oeuvre aux Sections 3.7 et 3.8 pour proposer des démonstrations "projectives" des Théorèmes de Pappus et de Desargues.
Si l'un de nos buts est de travailler en projectif pour démontrer des résultats affines, il nous faut tout d'abord nous représenter un espace affine comme "plongé" dans un espace projectif. Ce sera l'objet de la Section 3.5.

3.4.2 Equations cartésiennes de sous-espaces projectifs

Soit $F = p(\overrightarrow{F})$ un sous-espace projectif associé au sous-espace vectoriel $\overrightarrow{F}$. De façon naturelle :

Définition 3.6 *On appelle* ***système d'équations cartésiennes de*** F ***dans le repère projectif*** $(P_0, ..., P_{n+1}) = (p(e_0), ..., p(e_n), p(\sum_{i=0}^{n} e_i))$ *tout système d'équations cartésiennes de* $\overrightarrow{F}$ *dans la base* $(e_0, ..., e_n)$*.*

▶ Par exemple, une droite du plan projectif $\mathbb{P}^2(K)$ est formée de tous les points de coordonnées homogènes $(x : y : z)$ telles que

$$ax + by + cz = 0$$

où $(a, b, c) \in K^3 \backslash \{(0, 0, 0)\}$. On notera qu'une telle équation est homogène, ce qui est fort heureux puisque des coordonnées homogènes d'un point ne sont définies qu'à un coefficient multiplicatif non nul près.

▶ Deux points distincts A et B du plan projectif $\mathbb{P}^2(K)$ déterminent une droite que nous noterons (AB). Dans le repère projectif canonique de $\mathbb{P}^2(K)$ (associé à la base canonique de K^3), notons $(a_0 : a_1 : a_2)$ et $(b_0 : b_1 : b_2)$ les coordonnées homogènes respectives de A et B. Notons aussi a et b des vecteurs de K^3 tels que $A = p(a)$ et $B = p(b)$.
Un point $M = p(x)$, de coordonnées homogènes $(x_0 : x_1 : x_2)$, appartient à la droite (AB) si et seulement si le vecteur x appartient au plan vectoriel $\mathrm{Vect}(a, b)$ engendré par les vecteurs a et b. On peut donc écrire

$$M \in (AB) \;\Leftrightarrow\; x \in \mathrm{Vect}(a, b) \;\Leftrightarrow\; \begin{vmatrix} x_0 & a_0 & b_0 \\ x_1 & a_1 & b_1 \\ x_2 & a_2 & b_2 \end{vmatrix} = 0$$

et obtenir ainsi une équation de (AB).

Cette description de la droite projective (AB) nous permet d'énoncer :

Théorème 3.5 *Soient A, B, C, D quatre points du plan projectif $\mathbb{P}^2(K)$, de coordonnées homogènes respectives $(a_0 : a_1 : a_2)$, $(b_0 : b_1 : b_2)$, $(c_0 : c_1 : c_2)$ et $(d_0 : d_1 : d_2)$. Les propriétés suivantes sont équivalentes :*

i) (A, B, C, D) est un repère projectif de $\mathbb{P}^2(K)$,

ii) Trois points parmi A, B, C, D ne sont jamais alignés,

iii) Aucun mineur (i.e. déterminant d'une matrice extraite) d'ordre trois extrait de la matrice

$$\begin{pmatrix} a_0 & b_0 & c_0 & d_0 \\ a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \end{pmatrix}$$

n'est nul.

3.5 Lien affine-projectif

3.5.1 Les complémentaires d'hyperplans projectifs

Soit $\mathbb{P}^n = \mathbb{P}(K^{n+1})$ l'espace projectif standard de dimension n sur K.
Si $x \in \mathbb{P}^n$, notons $x = (x_0 : x_1 : ... : x_n)$ les coordonnées homogènes de x dans la base canonique de K^{n+1}. Soit H_i l'hyperplan d'équation $x_i = 0$ dans $\mathbb{P}^n$ et $A_i = \mathbb{P}^n \backslash H_i$ le complémentaire de cet hyperplan. On a

$$\begin{aligned} x \in A_i & \Leftrightarrow & x = (x_0 : x_1 : ... : x_n) \quad \text{avec } x_i \neq 0 \\ & \Leftrightarrow & x = (\frac{x_0}{x_i} : \frac{x_1}{x_i} : ... : 1 : ... : \frac{x_n}{x_i}) \end{aligned}$$

(où le 1 est à la i-ème position, $0 \leq i \leq n$). On peut donc définir la bijection

$$\begin{array}{rccc} f_i : & K^n & \rightarrow & A_i \\ & (x_0, x_1, ..., \widehat{x_i}, ..., x_n) & \mapsto & (x_0 : x_1 : ... : 1 : ... : x_n) \end{array}$$

où le chapeau au-dessus de x_i signifie que x_i est absent. La bijection réciproque s'écrit

$$\begin{array}{rccc} f_i^{-1} : & A_i & \rightarrow & K^n \\ & (x_0 : x_1 : ... : x_n) & \mapsto & (\frac{x_0}{x_i}, \frac{x_1}{x_i}, ..., \widehat{\frac{x_i}{x_i}}, ..., \frac{x_n}{x_i}). \end{array}$$

Théorème 3.6 *L'ensemble A_i est structuré en espace affine par transport de la structure affine de l'espace vectoriel K^n.*

Preuve : Rappelons qu'un ensemble E est structuré en espace affine sur K s'il existe un espace vectoriel $\overrightarrow{E}$ sur K ainsi qu'une application

$$\begin{array}{rccc}\varphi : & E\times\overrightarrow{E} & \to & E \\ & (M,\overrightarrow{u}) & \mapsto & M+\overrightarrow{u}\end{array}$$

telle que :

A1 $\forall M\in E\quad \forall\overrightarrow{u},\overrightarrow{v}\in\overrightarrow{E}\quad (M+\overrightarrow{u})+\overrightarrow{v}=M+(\overrightarrow{u}+\overrightarrow{v})$,

A2 Pour tout $M\in E$, l'application

$$\begin{array}{rccc}\varphi_M : & \overrightarrow{E} & \to & E \\ & \overrightarrow{u} & \mapsto & M+\overrightarrow{u}\end{array}$$

est bijective ([1], Définition 4)).

On peut vérifier que les conditions **A1** et **A2** entraînent $M+\overrightarrow{0}=M$ pour tout M, de sorte que dire que E est un espace affine revient à dire que le groupe additif $(\overrightarrow{E},+)$ d'un espace vectoriel opère simplement et transitivement sur E. Avec cette définition, il est facile de voir que l'espace vectoriel K^n est canoniquement structuré en espace affine par l'application $\varphi(x,y)=x+y$ (addition dans l'espace vectoriel K^n).

Le transport de structure de K^n vers A_i se matérialise lorsqu'on définit l'application

$$\begin{array}{rccc}\varphi : & A_i\times K^n & \to & A_i \\ & (a,x) & \mapsto & a+x=f_i\left(f_i^{-1}(a)+x\right),\end{array}$$

et que l'on vérifie que φ satisfait les axiomes **A1** et **A2**.

• **A1.** φ définit une opération de groupe car, pour tout $a\in A_i$ et pour tous $x,y\in K^n$,

$$\begin{aligned}(a+x)+y &= f_i\left(f_i^{-1}(a+x)+y\right)=f_i\left(f_i^{-1}\circ f_i\left(f_i^{-1}(a)+x\right)+y\right)\\ &= f_i\left(f_i^{-1}(a)+x+y\right)=a+(x+y),\end{aligned}$$

et $a+0=a$.

• **A2.** Cette opération est simplement transitive, autrement dit si $a,b\in A_i$ il existe un et un seul vecteur $x\in K^n$ tel que $a+x=b$. En effet,

$$a+x=b\;\Leftrightarrow\; f_i\left(f_i^{-1}(a)+x\right)=b\;\Leftrightarrow\; x=f_i^{-1}(b)-f_i^{-1}(a).\ \blacksquare$$

Visualisation : Si A_i' désigne l'hyperplan affine de K^{n+1} d'équation $x_i=1$, la bijection

$$\begin{array}{rccc}g_i : & A_i' & \to & A_i \\ & (x_0,x_1,...,1,...,x_n) & \mapsto & (x_0:x_1:...:1:...:x_n)\end{array}$$

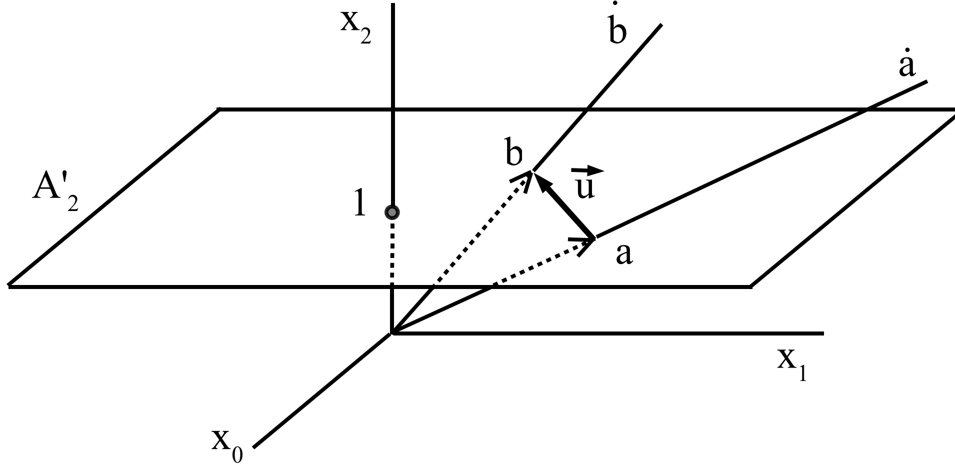


FIG. 3.3 – Visualisation d'une partie "affine" du plan projectif.

est le pendant naturel de la bijection $f_i : K^n \to A_i$, et permet aussi de définir la structure affine sur A_i.
La FIG. 3.3 permet de visualiser la situation lorsque $n = 2$. Elle montre deux représentants a et b (de cotes $x_2 = 1$ dans K^3) de deux points distincts $\dot{a}$, $\dot{b}$ de $A_2 = \mathbb{P}^2 \backslash H_2$, appartenant à l'hyperplan affine A'_2 de K^3 d'équation $x_2 = 1$, liés par la relation $a + \overrightarrow{u} = b$.
En effet, chaque point $\dot{a}$ de A_2 est identifié à un vecteur a de l'hyperplan affine A'_2, et l'action du groupe additif $(K^2, +)$ de l'espace vectoriel K^2 sur A_2 est donnée par

$$\forall a = (a_0 : a_1 : 1) \in A_2 \quad \forall \overrightarrow{u} = (\alpha, \beta) \in K^2 \quad a + \overrightarrow{u} = (a_0 + \alpha : a_1 + \beta : 1)\,.$$

Voici quelques conséquences du Théorème 3.6 :

Corollaire 3.1 *L'espace $\mathbb{P}^n$ est une variété différentielle de dimension n et de classe C^∞.*

Preuve : Les ensembles A_i recouvrent $\mathbb{P}^n$ et la famille $\{(A_i, f_i^{-1})_{0 \leq i \leq n}$ est un atlas C^∞ de $\mathbb{P}^n$. Les cartes sont les couples $\left(A_i, f_i^{-1}\right)$, et les changements de cartes $f_{ij} = f_j^{-1} \circ f_i$ sont de classe C^∞. Si $j > i$, on obtient :

$$\begin{array}{ccccc} f_i^{-1}(A_i \cap A_j) & \stackrel{f_i}{\to} & A_i \cap A_j & \stackrel{f_j^{-1}}{\to} & f_j^{-1}(A_i \cap A_j) \\ (x_0, ..., \widehat{x}_i, ..., x_n) & \mapsto & (x_0 : ... : 1_i : ... : x_n) & \mapsto & (\dfrac{x_0}{x_j}, ..., \dfrac{1_i}{x_j}, ..., \widehat{1}_j, ..., \dfrac{x_n}{x_j}). \blacksquare \end{array}$$

On dit que les ensembles A_i (on devrait dire : les couples $\left(A_i, f_i^{-1}\right)$) sont des **cartes affines** de $\mathbb{P}^n$. Tout point du projectif appartient à au moins l'une de ces cartes, si bien que, localement, un espace projectif est un espace affine.

Corollaire 3.2 *Le complémentaire d'un hyperplan dans un espace projectif de dimension n est un espace affine de dimension n.*

Preuve : Il suffit de choisir des coordonnées homogènes sur $\mathbb{P}^n$ bien adaptées à l'hyperplan projectif H, c'est-à-dire telles que H admette une équation de la forme $x_i = 0$, et utiliser le Théorème 3.6. ∎

Corollaire 3.3 *Tout espace affine E de dimension finie peut être plongé dans un espace projectif de même dimension. Il est alors isomorphe au complémentaire d'un hyperplan projectif.*

Preuve : Si E est un espace affine de dimension n, il est isomorphe à K^n par le choix d'un repère affine, puis en bijection avec l'une des parties $A_i = \mathbb{P}^n \backslash H_i$ via l'application $f_i : K^n \to A_i$. ∎

Définition 3.7 *Si le plongement est noté $f_i : E \simeq K^n \to A_i$, les éléments de $H_i = \mathbb{P}^n \backslash A_i$ sont appelés **points à l'infini** de E. L'hyperplan H_i est appelé **hyperplan à l'infini** de A_i, et parfois noté A_i^∞. Bien entendu :*

$$\mathbb{P}(K^n) = A_i \bigsqcup H_i = A_i \bigsqcup A_i^\infty.$$

Sur la FIG. 3.4 (où $n = 2$), la droite bleue D est dessinée dans le plan affine A_2'. La droite rouge D_∞ est le point à l'infini sur D, et la droite projective complète $(\dot{a}\dot{b})$ est la réunion : $(\dot{a}\dot{b}) = D \bigsqcup \{D_\infty\}$. On voit que les points à l'infini du plan affine A_2' d'équation $x_2 = 1$ sont les différentes directions des droites de A_2'. Si l'on identifie A_2' à A_2 et à K^2, alors

$$A_2' = A_2 = K^2 \subset \mathbb{P}(K^3)$$

et l'on peut écrire $\mathbb{P}(K^3) = A_2' \bigsqcup \mathbb{P}(A_2')$, ou encore $\mathbb{P}(K^3) = K^2 \bigsqcup \mathbb{P}\left(K^2\right)$, puisque $\mathbb{P}(A_2')$ est l'ensemble des droites de A_2'.

> Le plan projectif $\mathbb{P}(K^3)$ peut ainsi être considéré comme la réunion disjointe du plan affine K^2 et de l'ensemble de toutes les directions des droites de K^2.

Nous retrouvons la description donnée dans l'introduction, à la Section 3.1.

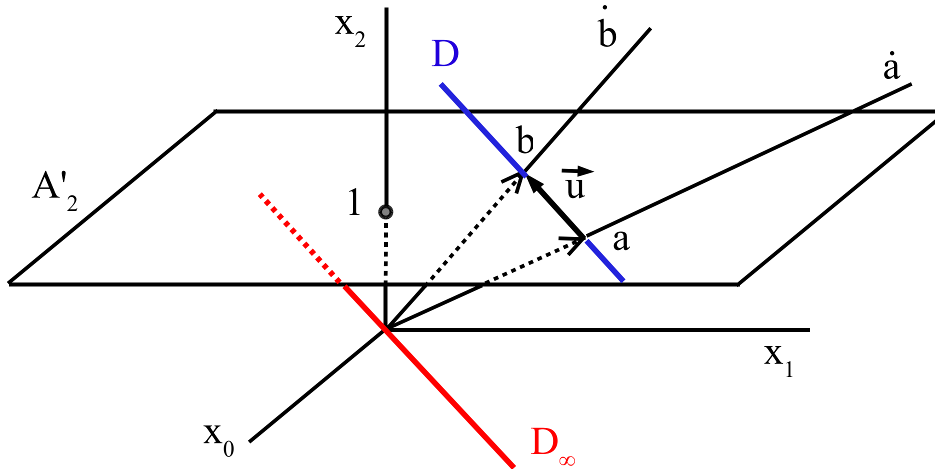


FIG. 3.4 – Une droite affine et son point à l'infini

3.5.2 Visualisation du plan projectif

Penser au plan projectif, c'est penser à un plan affine qui a été complété par une "droite à l'infini". Reprenons les notations de la Section précédente, mais supposons ici que $n = 2$. On peut compléter la FIG. 3.3 en dessinant un repère projectif (O, I, J, K) bien adapté à la carte affine A_2, c'est-à-dire formé de quatre points projectifs qui appartiennent à l'espace affine A_2, comme sur la FIG. 3.5. Le plus simple est de choisir les points O, I, J, K de coordonnées homogènes :

$$O\,(0:0:1)\ ;\quad I\,(1:0:1)\ ;\quad J\,(0:1:1)\ ;\quad K\,(1:1:1)\,.\qquad (*)$$

En fait, on peut procéder différemment en considérant quatre points quelconques O, I, J, K de A_2 tels que trois d'entre eux ne sont jamais alignés, rappeler qu'un tel système de points constitue un repère projectif de $\mathbb{P}(K^3)$, puis utiliser des coordonnées homogènes dans ce repère de façon à avoir encore $(*)$.

Chacun des points O, I, J, K possède des coordonnées homogènes, mais aussi des coordonnées affines puisqu'il s'agit de points de l'espace affine A_2 que l'on peut rapporter au repère affine (O, I, J). Les coordonnées affines des points O, I, J, K sont alors :

$$O\,(0,0)\ ;\quad I\,(1,0)\ ;\quad J\,(0,1)\ ;\quad K\,(1,1)\,,$$

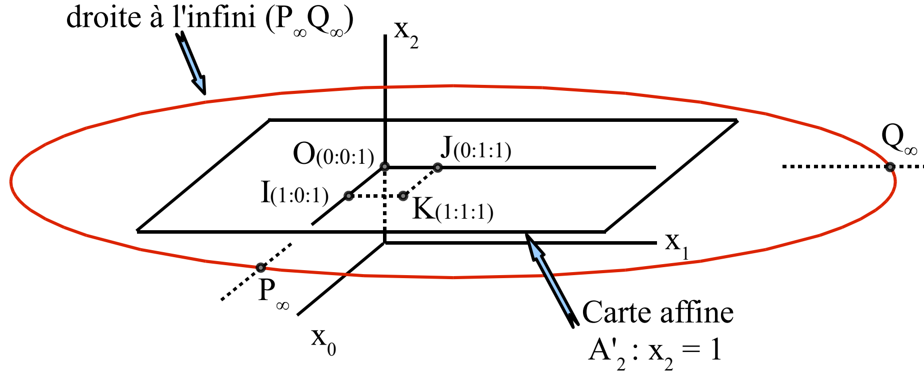


FIG. 3.5 – Un repère projectif (O, I, J, K) adapté au plan affine A_2.

où K est le point unitaire.

Les points P_∞ et Q_∞ de coordonnées homogènes respectives $(1:0:0)$ et $(0:1:0)$ n'appartiennent pas à A_i, mais plutôt à l'hyperplan à l'infini H_2 d'équation $x_2 = 0$. Ici, on parlera plutôt de "droite à l'infini" puisque $n = 2$ entraîne $\dim H_2 = n - 1 = 1$. On notera $H_2 = (P_\infty Q_\infty)$ cette droite.
Les points P_∞ et Q_∞ n'appartiennent pas à l'espace affine A_2, mais il peut être évocateur (et assez tentant) de leur trouver des coordonnées affines dans le plan A_2. Que pourrions-nous alors proposer ?

Compte tenu de l'identification de A_2 à K^2 donné par f_i^{-1} (à la Section 3.5.1), il est raisonnable (?) de dire que $P_\infty\,(1:0:0)$ correspond à un point affine de A_2 dont les coordonnées dans le repère affine (O, I, J) sont $\left(\frac{1}{0}, \frac{0}{0}\right)$. Hum, disons que $\frac{1}{0} = \infty$ et que $\frac{0}{0} = 0$, et l'on pourra dire que les coordonnées affines de P_∞ sont $(\infty, 0)$. C'est abusif, mais c'est parlant !
De même, on peut imaginer que les coordonnées affines de Q_∞ sont $(0, \infty)$.

Cela ne nous mène malheureusement pas bien loin car nous serions incapable de proposer des coordonnées affines "décentes" aux autres points de la droite $(P_\infty Q_\infty)$, à moins de proposer (∞, ∞), ce qui serait un comble[3].

En regardant la FIG. 3.5, il est tout de même agréable de penser que les trois points O, I, J formant un repère affine de A_2, le point unitaire K qui complète les trois points précédents pour en faire un repère projectif de $\mathbb{P}(K^3)$, et les points P_∞, Q_∞ qui déterminent la droite à l'infini de l'hyperplan A_2, ont des coordonnées homogènes dans $\mathbb{P}(K^3)$, et des coordonnées affines dans A_2 (avec l'abus signalé pour P_∞ et Q_∞) donnée par le tableau (T) ci-dessous.

[3]Une infinité de points de $(P_\infty Q_\infty)$ auraient alors les mêmes coordonnées !

Points	Coordonnées homogènes	Coordonnées affines
O	$(0:0:1)$	$(0,0)$
I	$(1:0:1)$	$(1,0)$
J	$(0:1:1)$	$(0,1)$
K	$(1:1:1)$	$(1,1)$
P_∞	$(1:0:0)$	$(\infty,0)$
Q_∞	$(0:1:0)$	$(0,\infty)$

Tableau (T)

On peut encore utiliser la FIG. 3.5 pour visualiser des droites parallèles de l'espace affine A_2 qui se coupent en un point de la droite à l'infini $(P_\infty Q_\infty)$. On obtient la FIG. 3.6 où l'on a dessiné trois droites parallèles du plan affine qui se coupent (et coupent la droite à l'infini) en M.
Vérifions-cela. De façon générale, toute droite affine D_c de A_2 de vecteur directeur donné $\overrightarrow{u}\,(-b,a)$ admet une équation de la forme

$$D_c : ax_0 + bx_1 + c = 0$$

(avec $(a,b,c) \in K$ et $(a,b) \neq (0,0)$) dans le repère affine (O,I,J).
Compte tenu du plongement de A_2 dans $\mathbb{P}(K^3)$, l'équation de l'unique droite projective $\widetilde{D}_c$ de $\mathbb{P}(K^3)$ qui contienne D_c sera

$$\widetilde{D}_c : ax_0 + bx_1 + cx_2 = 0.$$

Autrement dit, $\widetilde{D}_c$ est l'ensemble des points du projectif de coordonnées homogènes $(x_0 : x_1 : x_2)$ vérifiant $ax_0 + bx_1 + cx_2 = 0$, et $\widetilde{D}_c \cap A_2 = D_c$. Il s'agit maintenant de déterminer l'intersection $\widetilde{D}_c \cap (P_\infty Q_\infty)$. Si M est de coordonnées homogènes $(\alpha : \beta : \gamma)$, alors M appartient à $\widetilde{D}_c \cap (P_\infty Q_\infty)$ si et seulement si

$$\begin{cases} a\alpha + b\beta + c\gamma = 0 \\ \gamma = 0 \end{cases}$$

c'est-à-dire

$$\begin{cases} a\alpha = -b\beta \\ \gamma = 0, \end{cases}$$

et l'on peut affirmer que $\widetilde{D}_c \cap (P_\infty Q_\infty) = \{M\}$ où $M\,(-b : a : 0)$.

Cela nous montre deux choses :

- que des droites affines de A_2 parallèles entre elles se coupent toujours en un seul point sur la droite à l'infini $(P_\infty Q_\infty)$,

- que deux droites distinctes quelconques du projectif se coupent toujours en un point[4].

La FIG. 3.6 essaie de rendre compte de ce phénomène. On notera que, sur cette figure (comme sur la FIG. 3.5) la droite à l'infini $(P_\infty Q_\infty)$ est indûment représentée par un cercle, alors qu'on devrait plutôt dessiner un demi-cercle. Mais on peut s'affranchir de cela en imaginant qu'un point de $(P_\infty Q_\infty)$ est représenté par *deux points diamétralement opposés* sur ce cercle.

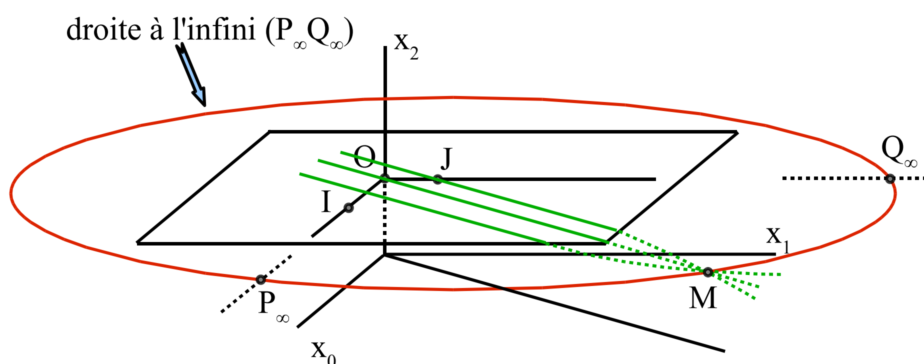


FIG. 3.6 – Droites parallèles de A_2 se coupant sur la droite à l'infini.

Une dernière remarque pour terminer cette Section. Nous avons-vu qu'il était toujours possible de choisir un bon repère projectif (O, I, J, K) pour lequel la carte affine A_2 admet le repère affine (O, I, J), et qu'il suffisait de choisir quatre points de A_2 tels que trois d'entre eux ne sont jamais alignés.
Mais étant donné deux points A et B de $\mathbb{P}(K^3)$, est-il toujours possible de trouver un repère projectif (O, I, J, K) pour lequel A et B sont des points "à l'infini", ce qui signifierait que, pour le choix d'un tel repère, la droite (AB) serait tout bonnement la droite à l'infini $(P_\infty Q_\infty)$ dont nous avons parlé plus haut ?
La réponse est affirmative, puisqu'il suffit de considérer que (AB) est la droite d'équation $x_2 = 0$, puis de choisir un repère projectif (O, I, J, K) formé de quatre points appartenant à $A_2 = \mathbb{P}(K^3)\setminus(AB)$, pour répondre à nos attentes. On dit que l'*on a envoyé les points A et B à l'infini*, ce qui sous-entend que l'on va s'intéresser aux points qui n'appartiennent pas à la droite (AB) et les considérer comme étant des points d'un espace affine où l'on peut dessiner, interpréter des configurations, etc.

[4] Nous le savions déjà : voir Théorème 3.3.

3.5.3 Visualisation dans le cas général

On reprend exactement le travail de la section précédente en nous plaçant cette fois-ci dans l'espace projectif $\mathbb{P}(K^{n+1})$ de dimension n. La carte affine $A_n = \mathbb{P}(K^{n+1})\backslash H_n$ où H_n désigne l'hyperplan projectif d'équation $x_n = 0$, est mise en valeur si l'on choisit un repère projectif

$$\mathcal{R} = (O, I_1, ..., I_n, K)$$

formé de points de A (tels que $n+1$ quelconques d'entre eux n'appartiennent pas à un même hyperplan). Le *bon choix* est sans doute de prendre :

$$\begin{cases} O = (0:0:...:0:0:1) = p\,(e_0) \\ I_1 = (1:0:...:0:0:1) = p\,(e_1) \\ .... \\ I_n = (0:0:...:0:1:1) = p\,(e_n) \\ K = (1:1:...:1:1:1) = p\,(e_0 + ... + e_n)\,. \end{cases}$$

On ne s'en privera pas dans la suite.

Dans l'écriture ci-dessus :

- on a identifié les points de $\mathbb{P}(K^{n+1})$ avec leurs coordonnées homogènes dans $\mathcal{R}$.

- p désigne la surjection canonique définie à la Section 3.2.

- le système $(e_0, ..., e_n)$ désigne une base de K^{n+1} associée au repère projectif $\mathcal{R}$ au sens de la Définition 3.5. On prendra bien garde de noter que $(e_0, ..., e_n)$ n'est pas la base canonique de K^{n+1}.

Avec ce choix, $\mathcal{R}_a = (O, I_1, ..., I_n)$ est un repère affine de A (muni de sa structure canonique d'espace affine de dimension n), et les coordonnées affines de l'origine O et des sommets I_1, ..., I_n de ce repère sont données dans le tableau (T_n) :

Points	Coordonnées homogènes dans $\mathcal{R}$	Coordonnées affines dans $\mathcal{R}_a$
O	$(0:0:...:0:0:1)$	$(0,0,...,0,0)$
I_1	$(1:0:...:0:0:1)$	$(1,0,...,0,0)$
$\vdots$	$\vdots$	$\vdots$
I_n	$(0:0:...:0:1:1)$	$(0,0,...,0,1)$
K	$(1:1:...:1:1:1)$	$(1,1,...,1,1)$

Tableau (T_n)

3.6 Topologie sur $\mathbb{P}(E)$

Soient n un entier naturel non nul, et E un espace préhilbertien de dimension finie $n+1$ sur le corps K (où $K = \mathbb{R}$ ou $\mathbb{C}$). La norme de E est notée $|| \ ||$ et l'on pose $U = \{z \in K \ / \ |z| = 1\}$. On note $\mathbb{P}(E)$ l'espace projectif (de dimension n) associé à E, et $p : E \to \mathbb{P}(E)$ la surjection canonique qui à un vecteur non nul associe la droite qu'il dirige.

3.6.1 Une distance sur $\mathbb{P}(E)$

Théorème 3.7 *Si $x, y \in \mathbb{P}(E)$, et si u, v désignent des vecteurs normés de E tels que $x = p(u)$ et $y = p(y)$, l'application $d : \mathbb{P}(E) \times \mathbb{P}(E) \to \mathbb{R}_+$ qui au couple (x, y) associe le réel $d(x, y) = \text{Min}_{(\lambda,\mu)\in U^2} ||\lambda u - \mu v||$ est une distance sur $\mathbb{P}(E)$.*

Preuve : On remarque d'abord que :

▷ Le nombre $d(x, y)$ proposé est indépendant du choix du couple (u, v) tel que $x = p(u)$ et $y = p(y)$,

▷ Le minimum existe bel et bien, puisque l'application $(\lambda, \mu) \mapsto ||\lambda u - \mu v||$, continue sur le compact U^2, atteint ses bornes sur ce compact.

▷ $d(x, y) = \text{Min}_{\lambda\in U} ||\lambda u - v||$.

On vérifie ensuite les trois axiomes d'une distance :

• Si $d(x, y) = 0$, en notant $d(x, y) = ||\lambda_0 u - \mu_0 v||$ où λ_0 et μ_0 sont des éléments de U, on obtient $\lambda_0 u - \mu_0 v = 0$, d'où $x = y$.

• Pour tous $x, y \in \mathbb{P}(E)$, $d(x, y) = d(y, x)$.

• Si $z = p(w) \in \mathbb{P}(E)$,

$$\forall \lambda, \mu \in U \quad ||\lambda u - \mu v|| \leq ||\lambda u - w|| + ||w - \mu v||$$

donc

$$\forall \lambda, \mu \in U \quad d(x, y) \leq ||\lambda u - w|| + ||w - \mu v||.$$

Pour μ fixé, il suffit de passer à la borne inférieure pour $\lambda \in U$ dans l'inégalité $d(x, y) - ||w - \mu v|| \leq ||\lambda u - w||$ pour obtenir

$$\forall \mu \in U \quad d(x, y) \leq d(x, z) + ||w - \mu v||.$$

On passe ensuite à la borne inférieure dans $d(x, y) - d(x, z) \leq ||w - \mu v||$ pour $\mu \in U$, pour obtenir $d(x, y) \leq d(x, z) + d(z, y)$. ■

Théorème 3.8 *Muni de la distance d définie au Théorème 3.7, $\mathbb{P}(E)$ est un espace métrique compact et la surjection canonique $p : E \to \mathbb{P}(E)$ est continue.*

Preuve : L'application $p : E \to \mathbb{P}(E)$ est lipschitzienne de rapport 1 puisque

$$\forall u, v \in E \quad d(p(u), p(v)) = \underset{(\lambda,\mu)\in U^2}{\mathrm{Min}} ||\lambda u - \mu v|| \leq ||u - v||.$$

C'est donc une application continue. L'espace $\mathbb{P}(E)$ est compact comme l'image de la sphère (compacte) $S(E) = \{x \in E \,/\, ||x|| = 1\}$ par l'application continue p. ■

Puisqu'un espace métrique compact est nécessairement complet (on vérifiera en effet en exercice que toute suite de Cauchy qui possède au moins une valeur d'adhérence l est convergente vers l), on peut affirmer que $\mathbb{P}(E)$ est complet.

E étant identifié à K^{n+1} par le choix d'une base, si $x \in \mathbb{P}(E)$, on note $x = (x_0 : x_1 : ... : x_n)$ les coordonnées homogènes de x dans cette base, H_i l'hyperplan d'équation $x_i = 0$ dans $\mathbb{P}(E)$ et $A_i = \mathbb{P}(E) \backslash H_i$ l'ouvert affine, complémentaire de H_i dans $\mathbb{P}(E)$.

Théorème 3.9 *Les isomorphismes affines $f_i : K^n \to A_i$ sont des homéomorphismes, et $\{A_i\}_{i=0,...,n}$ est un recouvrement de $\mathbb{P}(E)$ par des ouverts partout denses.*

Preuve : ▶ Montrons que l'application

$$\begin{array}{cccc} f_i : & K^n & \to & A_i = \mathbb{P}(E) \backslash H_i \\ & (x_0, x_1, ..., \widehat{x_i}, ..., x_n) & \mapsto & (x_0 : x_1 : ... : 1 : ... : x_n) \end{array}$$

est un homéomorphisme. Pour cela, décomposons f_i en $f_i = p' \circ r \circ q$ où

$$\begin{array}{cccc} q : & K^n & \to & A'_i = K^{n+1} \cap \{x_i = 1\} \\ & (x_1, ..., x_n) & \mapsto & (x_1, ..., x_{i-1}, 1, x_{i+1}, ..., x_n) \end{array}$$

est une bijection bicontinue,

$$\begin{array}{cccc} r : & A'_i & \to & S_i^+ \\ & y & \mapsto & y/||y|| \end{array}$$

est une bijection bicontinue de A'_i sur $S_i^+ = S^n(K) \cap \left\{(x_0, ..., x_n) \,/\, x_i \in \mathbb{R}_+^*\right\}$, où l'on pose $S^n(K) = \left\{x \in K^{n+1} \,/\, ||x|| = 1\right\}$, de bijection réciproque

$$\begin{array}{cccc} r^{-1} : & S_i^+ & \to & A'_i \\ & (x_0, ..., x_n) & \mapsto & (\frac{x_0}{x_i}, ..., \frac{x_n}{x_i}), \end{array}$$

et où $p' = p|_{S_i^+} : S_i^+ \to A_i$ désigne la restriction (au départ et à l'arrivée) de la projection canonique $p : E \to \mathbb{P}(E)$ à S_i^+.

Puisque q et r sont déjà des bijections bicontinues, montrer que f_i est un homéomorphisme revient à montrer que p' est une bijection bicontinue.
On remarque tout d'abord que p' est bijective et continue (comme restriction de l'application p, continue d'après le Théorème 3.8). Il reste seulement à vérifier que p'^{-1} est continue, autrement dit p' est une application fermée (c'est-à-dire transforme tout fermé de S_i^+ en un fermé de A_i).
On remarque que l'ensemble de départ S_i^+ est un sous-espace métrique du compact $S_i^+ \cup S_i^0$ où $S_i^0 = S^n(K) \cap \{(x_0, x_1, ..., x_n) \,/\, x_i = 0\}$ (l'adhérence de S_i^+ est égale à $S_i^+ \cup S_i^0$). Tout fermé F' de S_i^+ est donc la trace d'un fermé F de $S_i^+ \cup S_i^0$, soit $F' = F \cap S_i^+$. Mais F, fermé dans un compact, est compact. On a

$$p\left(F'\right) = p(F) \cap p\left(S_i^+\right) = p(F) \cap A_i,$$

et $p(F')$ apparaît comme l'intersection de A_i et du compact $p(F)$ (image du compact F par p continue). $p(F')$ est donc, par définition, un fermé de A_i.

▶ Il est évident que la famille $\{A_i\}_{i=0,...,n}$ recouvre $\mathbb{P}(E)$.

▶ Montrer que $A_i = \mathbb{P}(E) \setminus H_i = \mathbb{P}(E) \setminus p(S_i^0)$ est ouvert dans $\mathbb{P}(E)$ revient à prouver que $p(S_i^0)$ est fermé, ce qui n'est pas difficile puisque $p(S_i^0)$ est compact comme l'image du compact S_i^0 par l'application continue p.

▶ Il reste à montrer que chaque ouvert A_i est dense dans $\mathbb{P}(E)$. Pour cela, il s'agit de prouver que tous les points $x = p(u)$ de $\mathbb{P}(E) \setminus A_i$ (où $u \in K^{n+1}$) appartiennent à l'adhérence de A_i. Notons $u = (x_0, ..., x_{i-1}, 0, x_{i+1}, ..., x_n)$ et définissons la suite $(x_m)_{m \in \mathbb{N}^*}$ par $x_m = p(u_m)$ et

$$u_m = (x_0, ..., x_{i-1}, \frac{1}{m}, x_{i+1}, ..., x_n)$$

pour tout $m \in \mathbb{N}^*$. Puisque $\lim u_m = u$, la continuité de p montre que

$$\lim x_m = \lim p(u_m) = p(u) = x,$$

et il suffit de remarquer que x_m appartient à A_i pour tout m pour conclure. ■

3.6.2 Points à l'infini

On conserve les notations de la Section 3.6.1. La frontière ∂A_i de A_i est l'ensemble

$$\partial A_i = \overline{A}_i \setminus \overset{o}{A}_i = \mathbb{P}(E) \setminus A_i$$

que nous noterons A_i^∞. C'est le complémentaire de A_i dans $\mathbb{P}(E)$. On rappelle que (Section 3.5.1) :

Définition 3.8 *Par définition,* $A_i^\infty = \mathbb{P}(E)\backslash A_i = \partial A_i$ *est* ***l'ensemble des points à l'infini de*** A_i*. C'est aussi l'ensemble des directions des droites de* $A_i' = K^{n+1} \cap \{x_i = 1\}$ *que l'on confondra avec* A_i*.*

Avec cette définition, $\mathbb{P}(E)$ apparaît comme la réunion disjointe de A_i et des directions des droites de A_i, ce que l'on note

$$\mathbb{P}(E) = A_i \cup A_i^\infty.$$

Cette description de $\mathbb{P}(E)$ comme réunion d'un espace affine A_i et d'un ensemble de points A_i^∞ formés de "points à l'infinis" de A_i, est précieuse et "parle à notre imagination". Les points de A_i^∞ complètent idéalement notre espace affine A_i en en faisant un espace métrique compact.
On fera attention tout de même : cette "complétion" n'est parfaite que pour la métrique que nous avons définie sur $\mathbb{P}(E)$, et non pour la distance hilbertienne usuelle sur A_i, puisque, pour cette distance hilbertienne, A_i est déjà un espace complet !
L'expression "points à l'infini" est justifiée par le Théorème suivant :

Théorème 3.10 *Pour tout suite* (x_k) *de l'espace affine* A_i*, et en notant* $||x_k||$ *la norme euclidienne de* x_k *dans* A_i*,*

$$\lim_{k\to+\infty} d(x_k, A_i^\infty) = 0 \;\Leftrightarrow\; \lim_{k\to+\infty} ||x_k|| = +\infty.$$

Preuve : Conservons les notations de la preuve du Théorème 3.9. Par définition de la distance d dans $\mathbb{P}(E)$,

$$d(x_k, A_i^\infty) = d\left(\frac{x_k}{||x_k||}, S_i^0\right).$$

Le Théorème de Pythagore appliqué deux fois montre que pour un point quelconque $y = (y_0..., y_n)$ de la sphère $S(E)$,

$$d\left(y, S_i^0\right) = \sqrt{2 - 2\sqrt{1 - |y_i^2|}} := \varphi\left(|y_i^2|\right).$$

On trouve en effet (en utilisant la FIG. 3.7 qui représente la situation en coupe) :

$$\begin{aligned} d\left(y, S_i^0\right) = MP &= \sqrt{MH^2 + HP^2} \\ &= \sqrt{|y_i^2| + (\overline{OP} - \overline{OH})^2} \\ &= \sqrt{|y_i^2| + (1 - \sqrt{1 - |y_i^2|})^2} = \sqrt{2 - 2\sqrt{1 - |y_i^2|}}. \end{aligned}$$

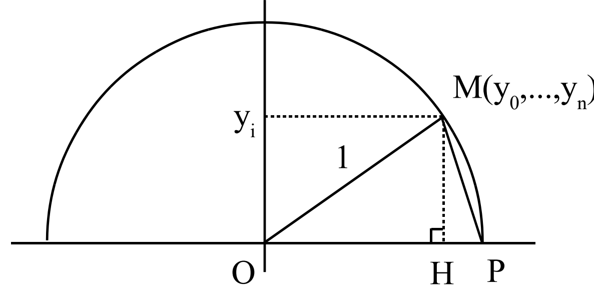


FIG. 3.7 – Calcul de $d\left(y, S_i^0\right) = MP$

Le point $x_k \in A_i$ de coordonnées homogènes $(x_0^{(k)}; ...; x_{i-1}^{(k)}; 1; x_{i+1}^{(k)}; ...; x_n^{(k)})$ est identifié au point $(x_0^{(k)}, ..., x_{i-1}^{(k)}, x_{i+1}^{(k)}, ..., x_n^{(k)})$ de l'espace affine A_i. Ainsi $||x_k|| = \sqrt{(x_0^{(k)})^2 + ... + (x_{i-1}^{(k)})^2 + (x_{i+1}^{(k)})^2 + ... + (x_n^{(k)})^2}$, et la i-ème coordonnée de $x_k/||x_k||$ est $1/||x_k||$. On a donc

$$d(x_k, A_i^\infty) = d\left(\frac{x_k}{||x_k||}, S_i^0\right) = \varphi\left(\frac{1}{||x_k||^2}\right),$$

et l'on peut écrire

$$||x_k|| \to +\infty \quad \Leftrightarrow \quad \frac{1}{||x_k||^2} \to 0 \quad \Leftrightarrow \quad d(x_k, A_i^\infty) \to 0. \blacksquare$$

3.6.3 Lien avec la topologie-quotient

De façon générale, si E est un espace topologique et si $f : E \to F$ est une application de E dans un ensemble F, on appelle topologie finale sur F (pour f) la topologie la plus fine rendant l'application f continue. Cela revient au même de dire qu'une partie U de F est un ouvert pour cette topologie si et seulement si $f^{-1}(U)$ est un ouvert de E.
Ici, nous disposons de la surjection canonique

$$p : E \to \mathbb{P}(E) = (E \setminus \{0\})/\mathcal{R}$$

où la relation d'équivalence $\mathcal{R}$ est définie par

$$x\mathcal{R}y \Leftrightarrow \exists \lambda \in K^* \quad x = \lambda y.$$

La topologie finale sur $\mathbb{P}(E)$, appelée topologie-quotient, est la topologie la plus fine rendant la surjection canonique p continue. Le résultat important est :

Théorème 3.11 *Sur* $\mathbb{P}(E) = (E\backslash\{0\})/\mathcal{R}$*, la topologie induite par la distance* d *est égale à la topologie-quotient.*

Preuve : Notons :

$\mathcal{O}_q$ la topologie-quotient sur $\mathbb{P}(E)$. Une partie U de $\mathbb{P}(E)$ appartient à $\mathcal{O}_q$ si et seulement si $p^{-1}(U)$ est un ouvert de $E\backslash\{0\}$.

$\mathcal{O}_m$ la topologie de l'espace métrique $\mathbb{P}(E)$ induite par la distance d définie au Théorème 3.7.

Puisque $p : E\backslash\{0\} \to \mathbb{P}(E)$ est continue (pour l'espace métrique $\mathbb{P}(E)$, Théorème 3.8), toute image réciproque $p^{-1}(B(x,\varepsilon))$ d'une boule ouverte $B(x,\varepsilon)$ de centre $x \in \mathbb{P}(E)$ et de rayon ε, sera un ouvert de $E\backslash\{0\}$, et par conséquent toute boule ouverte $B(x,\varepsilon)$ sera un ouvert pour la topologie $\mathcal{O}_q$. On a montré l'inclusion

$$\mathcal{O}_m \subset \mathcal{O}_q.$$

Réciproquement, si $U \in \mathcal{O}_q$, alors $p^{-1}(U)$ est un ouvert de $E\backslash\{0\}$, donc s'écrit comme une réunion de boules ouvertes de E, disons

$$p^{-1}(U) = \bigcup_{i\in I} B(u_i, \varepsilon_i).$$

Alors

$$U = p(p^{-1}(U)) = \bigcup_{i\in I} p(B(u_i,\varepsilon_i)),$$

et pour montrer que U appartient à $\mathcal{O}_m$, il suffit de prouver que $p(B(u_i,\varepsilon_i))$ appartient à $\mathcal{O}_m$ pour tout i. Cela prouvera l'inclusion $\mathcal{O}_q \subset \mathcal{O}_m$, et l'on pourra conclure à l'égalité $\mathcal{O}_m = \mathcal{O}_q$.
Montrons donc que

$$\forall u \in E\backslash\{0\} \quad \forall \varepsilon \in \mathbb{R}_+^* \qquad p(B(u,\varepsilon)) \in \mathcal{O}_m.$$

Soit $x_0 = p(u_0) \in p(B(u,\varepsilon))$, avec $u_0 \in B(u,\varepsilon)$. Pour tout $x' = p(u') \in \mathbb{P}(E)$, il existe $\mu_0 \in U$ tel que

$$d(x', x_0) = \operatorname*{Min}_{\mu\in U} ||\mu u' - u_0|| = ||\mu_0 u' - u_0||.$$

On cherche un réel $\eta > 0$ tel que $B(x_0,\eta)$ soit incluse dans $p(B(u,\varepsilon))$, autrement dit tel que

$$d(x', x_0) < \eta \Rightarrow (x' = p(u') \text{ et } u' \in B(u,\varepsilon)). \quad (*)$$

Il existe un scalaire $\lambda_0 \in U$ tel que $d(x', x) = ||\lambda_0\mu_0 u' - u||$. On a

$$||\lambda_0\mu_0 u' - u|| \leq ||\lambda_0\mu_0 u' - \lambda_0 u_0|| + ||\lambda_0 u_0 - u||$$

donc il suffit d'avoir

$$||\lambda_0\mu_0 u' - \lambda_0 u_0|| < \varepsilon - ||\lambda_0 u_0 - u||$$

ou encore

$$||\mu_0 u' - u_0|| < \frac{\varepsilon - ||\lambda_0 u_0 - u||}{|\lambda_0|} := \eta$$

pour obtenir $||\mu_0 u' - u|| < \varepsilon$. On vient de prouver l'implication $(*)$, plus précisément

$$d\left(x', x_0\right) < \eta \Rightarrow \left(x' = p(\mu_0 u') \text{ et } \mu_0 u' \in B\left(u, \varepsilon\right)\right). \blacksquare$$

3.7 Le Théorème de Pappus

Nous nous plaçons dans la configuration de la FIG. 3.8.

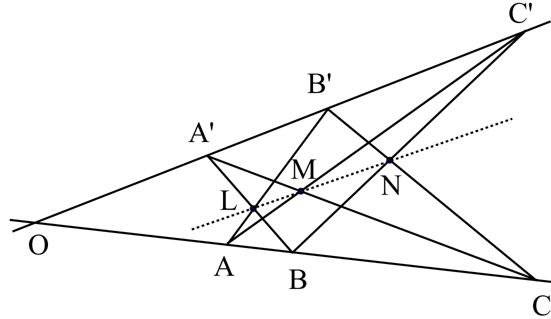


FIG. 3.8 – Configuration de Pappus (2 droites et 2 × 3 points)

Théorème 3.12 *On considère deux droites D et D' du plan projectif, et trois points distincts sur chacune de ces droites : A, B, C sur D, et A', B', C' sur D'. Les intersections L de (AB') et $(A'B)$, M de (AC') et $(A'C)$, et N de $(B'C)$ et (BC') sont des points alignés.*

Preuve : Les quatre points A, N, A', C sont trois à trois non alignés, si bien qu'on peut les choisir comme repère projectif (A, N, A', C). Les coordonnées homogènes de ces quatre points sont donc $A\,(1:0:0)$, $N\,(0:1:0)$, $A'\,(0:0:1)$, $C\,(1:1:1)$. Cela nous offre les équations de trois droites :

$$(CA) : y = z \; ; \quad (CN) : x = z \; ; \quad \left(CA'\right) : x = y.$$

Ces trois droites contiennent respectivement B, B' et M, donc il existe des scalaires p, q, r tels que $B\,(p:1:1)$, $B'\,(1:q:1)$, $M\,(1:1:r)$. Cela nous permet d'écrire les équations de 6 autres droites :

$$\begin{array}{rcl} (A'B') & : & y = qx \ ; \quad (AM) : z = ry \ ; \quad (BN) : x = pz \ ; \\ (A'B) & : & x = py \ ; \quad \big(AB'\big) : y = qz \ ; \quad (MN) : z = rx. \end{array}$$

Puisque $L \in (A'B) \cap (AB')$, ses coordonnées homogènes $(x:y:z)$ vérifient

$$\left\{ \begin{array}{l} x = py \\ y = qz, \end{array} \right.$$

et montrer que L appartient à (MN) revient à prouver l'implication

$$\left\{ \begin{array}{l} x = py \\ y = qz \end{array} \right. \Rightarrow \ z = rx. \quad (*)$$

On remarque que C' est le point de concours des trois droites $(A'B')$, (AM) et (BN). Ses coordonnées vérifient donc

$$\left\{ \begin{array}{l} y = qx \\ z = ry \\ x = pz \end{array} \right.$$

d'où $xyz = prqxyz$. On a $xyz \neq 0$ (en effet, si l'une des coordonnées x, y ou z est nulle, alors toutes les coordonnées de C' sont nulles, ce qui est impossible), et l'on obtient donc la relation $pqr = 1$. C'est cette relation qui nous permet de vérifier l'implication $(*)$. En effet

$$\left\{ \begin{array}{l} x = py \\ y = qz \end{array} \right. \Rightarrow \ x = pqz \ \Rightarrow \ rx = pqrz \ \Rightarrow \ z = rx. \ \blacksquare$$

Remarques : 1) Sur la FIG. 3.9, on a reporté les coordonnées homogènes des points du repère projectif dans lequel on travaille, ainsi que les équations de trois droites.

2) Le recours au projectif nous a permis de donner une seule démonstration pour des figures "affines" bien différentes. Les dessins suivant montrent des situations fort distinctes qui entrent néanmoins dans le cadre général d'une configuration de Pappus. En appelant

$$\mathcal{T} = \big\{((A'B),(AB')),((A'C),(AC')),((B'C),(BC'))\big\}$$

l'ensemble des trois couples de droites qui interviennent dans Pappus, on obtient 6 cas de figures :

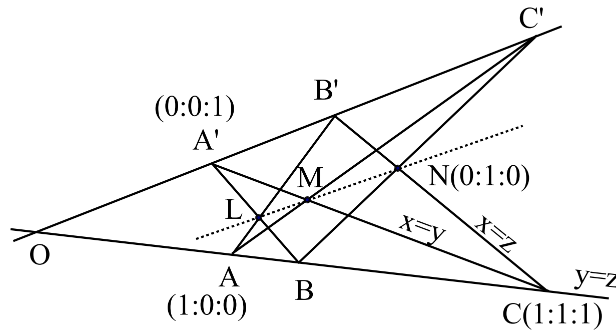


FIG. 3.9 – Choix du repère projectif

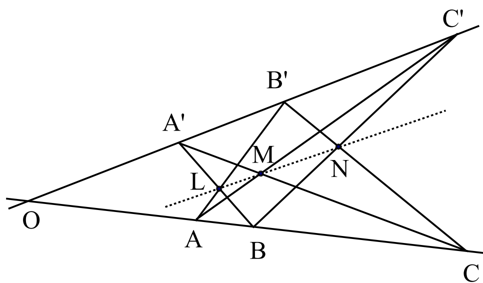
▷ Si D et D' sont sécantes,

- Cas n°1 : Tous les couples de $\mathcal{T}$ sont formés de droites sécantes,
- Cas n°2 : Seulement deux couples de $\mathcal{T}$ sont formés de droites sécantes,
- Cas n°3 : Tous les couples de $\mathcal{T}$ sont formés de droites parallèles.

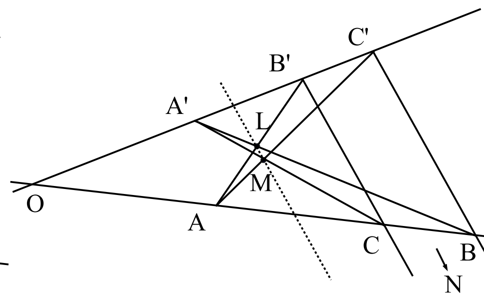
▷ Si D et D' sont parallèles,

- Cas n°4 : Tous les couples de $\mathcal{T}$ sont formés de droites sécantes,
- Cas n°5 : Seulement deux couples de $\mathcal{T}$ sont formés de droites sécantes,
- Cas n°6 : Tous les couples de $\mathcal{T}$ sont formés de droites parallèles.

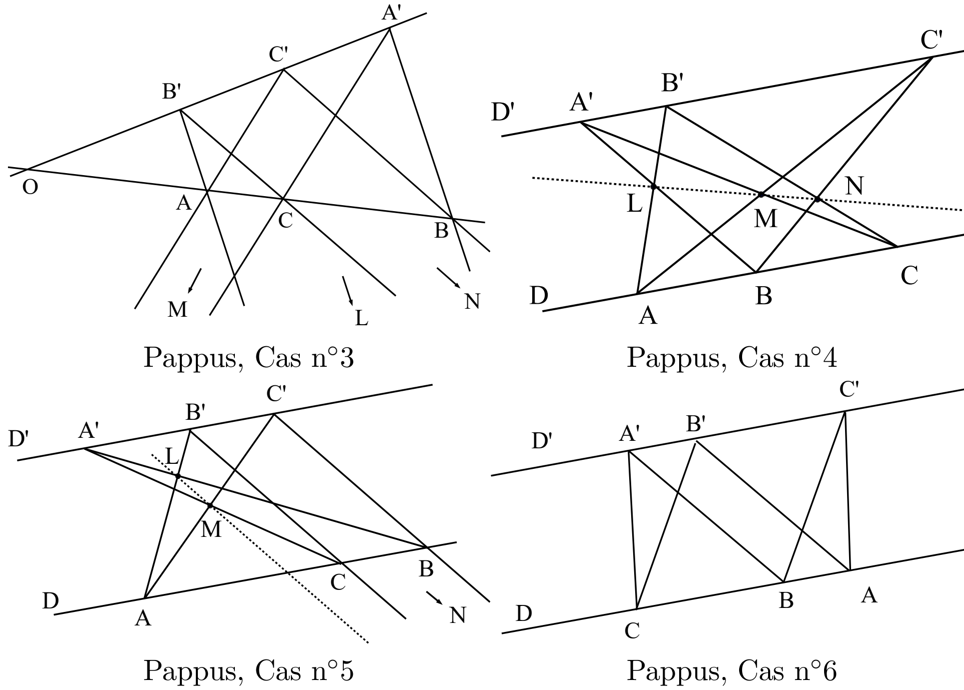
3) Une preuve affine du Théorème de Pappus dans le Cas n°1, et utilisant le Théorème de Ménélaüs, est donnée en ([1], Théorème 66). Le Cas n°6, qui constitue une application intéressante des propriétés des homothéties-translations, est traité en ([2], Théorème 63).



Pappus, Cas n°1



Pappus, Cas n°2



Pappus, Cas n°3 Pappus, Cas n°4

Pappus, Cas n°5 Pappus, Cas n°6

3.8 Le Théorème de Desargues

3.8.1 Le Théorème et sa preuve projective

La configuration de Desargues est formée de trois droites concourantes ou parallèles contenant chacune deux points.
L'étude des cas particuliers (existence ou non de points d'intersection de deux droites du dessin) complique la démonstration du Théorème dans le cadre affine, mais n'est pas nécessaire dans le cadre projectif où deux droites projectives s'interceptent inévitablement.
La simplicité de la preuve, effectuée une fois pour tous les cas de figure, illustre l'avantage qui existe à travailler dans le plan projectif.

Théorème 3.13 *Dans le plan projectif, on considère trois droites distinctes et concourantes D, D', D'', et six points distincts placés sur chacune de ces droites : A, B sur D ; A', B' sur D' ; et A'', B'' sur D''. Si L, M et N désignent les points d'intersection des couples de droites* $((AA'),(BB'))$, $((AA''),(BB''))$ *et* $(A'A''),(B'B'')$, *alors les points L, M, N sont alignés.*

Preuve : On peut supposer que le triangle $AA'A''$ n'est pas aplati, autrement l'alignement des points L, M, N est trivial sur la droite $(AA'A'')$. Notons O le point commun aux trois droites D, D', D'' et raisonnons sur la

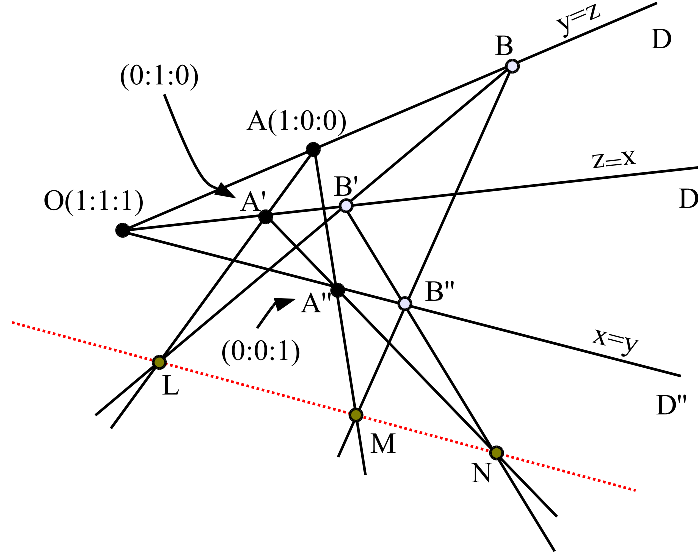


FIG. 3.10 – Configuration de Desargues

FIG. 3.10. Plaçons-nous dans le repère projectif formé par les points A, A', A'', O (trois points quelconques choisis parmi ces quatre points ne sont jamais alignés). Les coordonnées homogènes de ces points sont $A\,(1:0:0)$, $A'\,(0:1:0)$, $A''\,(0:0:1)$, $O\,(1:1:1)$, et l'on trouve facilement les équations des 6 droites suivantes :

$$\begin{array}{lll} D = (OA) : y = z \ ; & D' = \big(OA'\big) : x = z \ ; & D'' = \big(OA''\big) : x = y \ ; \\ \big(AA'\big) : z = 0 \ ; & \big(AA''\big) : y = 0 \ ; & \big(A'A''\big) : x = 0. \end{array}$$

Puisque $B \in (OA)$, $B' \in (OA')$ et $B'' \in (OA'')$, il existe trois scalaires p, q, r tels que $B\,(p:1:1)$, $B'\,(1:q:1)$ et $B''\,(1:1:r)$. Notre travail est maintenant clair : on va déterminer les coordonnées homogènes des points L, M et N, puis vérifier que ces points sont alignés.

- Cherchons les coordonnées homogènes $(x:y:z)$ de L.

On a $L \in (AA') \cap (BB')$, et une équation projective de (BB') est

$$\begin{vmatrix} x & p & 1 \\ y & 1 & q \\ z & 1 & 1 \end{vmatrix} = 0$$

soit $(BB') : (1-q)\,x - (p-1)\,y + (pq-1)\,z = 0$. Les coordonnées de L vérifient

donc le système

$$\begin{cases} z = 0 \\ (1-q)\,x + (1-p)\,y + (pq-1)\,z = 0 \end{cases}$$

donc $L\,(p-1 : 1-q : 0)$.

• De la même façon $M \in (AA'') \cap (BB'')$, donc M appartient à la droite (AA'') d'équation $y = 0$ et à la droite (BB'') d'équation

$$\begin{vmatrix} x & p & 1 \\ y & 1 & 1 \\ z & 1 & r \end{vmatrix} = (r-1)\,x - (pr-1)\,y + (p-1)\,z = 0.$$

Ainsi $M\,(1-p : 0 : r-1)$.

• Enfin $N \in (A'A'') \cap (B'B'')$, donc N appartient à la droite $(A'A'')$ d'équation $x = 0$ et à la droite $(B'B'')$ d'équation

$$\begin{vmatrix} x & 1 & 1 \\ y & q & 1 \\ z & 1 & r \end{vmatrix} = (qr-1)\,x - (r-1)\,y + (1-q)\,z = 0.$$

Ainsi $N\,(0 : 1-q : r-1)$.

• Montrer que L, M et N sont alignés revient à montrer que le déterminant dont les colonnes sont formées des coordonnées homogènes de L, M et N est nul. La vérification est aisée :

$$\begin{vmatrix} p-1 & 1-p & 0 \\ 1-q & 0 & 1-q \\ 0 & r-1 & r-1 \end{vmatrix} = (p-1)\,(1-q)\,(r-1) \begin{vmatrix} 1 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{vmatrix} = 0. \blacksquare$$

Remarques : 1) Dans la dernière partie de la preuve, on utilise une caractérisation de l'alignement des points L, M et N utilisant le déterminant des coordonnées homogènes de ces points. Cette caractérisation provient des définitions mêmes des droites projectives, images de plans vectoriels d'un espace vectoriel E de dimension 3 par la projection canonique $p : E \backslash \{0\} \to \mathbb{P}\,(E)$. Si l, m et n désignent des vecteurs de E qui se projettent respectivement sur L, M et N par p, autrement dit si $L = p\,(l)$, $M = p\,(m)$ et $N = p\,(n)$, alors

$$L,\, M,\, N \text{ alignés } \Leftrightarrow\ l,\, m,\, n \text{ coplanaires } \Leftrightarrow \begin{vmatrix} x_l & x_m & x_n \\ y_l & y_m & y_n \\ z_l & z_m & z_n \end{vmatrix} = 0$$

où (x_l, y_l, z_l), (x_m, y_m, z_m), (x_n, y_n, z_n) représentent les coordonnées respectives des vecteurs l, m, n dans une base (e_1, e_2, e_3) de E. Par définition, les triplets $(x_l : y_l : z_l)$, $(x_m : y_m : z_m)$, $(x_n : y_n : z_n)$ sont aussi les coordonnées homogènes des points L, M et N dans cette base.

2) L'homographie transformant les points A, B, C respectivement en A', B', C' permet d'obtenir le Théorème de Pappus comme une conséquence du Théorème de Desargues ([3], §. 2.2.2).

3) La réciproque du Théorème 3.13 est vraie ([3], Th. 2.4) : Si les points L, M et N sont alignés, alors les droites D, D', D'' sont concourantes.

3.8.2 Les différentes figures possibles

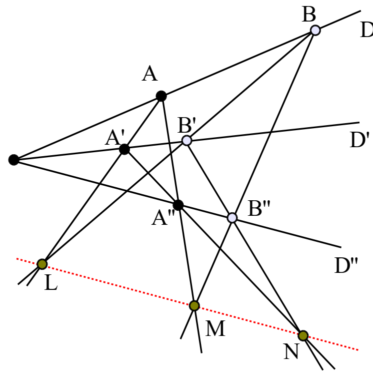
On peut dessiner les différentes figures que l'on peut obtenir. En géométrie affine, une démonstration particulière serait à donner pour chacune de ces figures.
L'utilisation du plan projectif permet à la fois de simplifier l'énoncé du résultat et d'unifier les différentes preuves correspondant à des cas de figures bien différentes. Finalement, en une seule preuve, nous avons démontré "beaucoup".
On notera que, du point de vue projectif, il n'existe pas plusieurs Théorème de Desargues, mais un seul énoncé qui possède de multiples facettes affines. Si l'on note

$$\mathcal{T} = \{((AA'), (BB')), ((BB'), (CC')), ((CC'), (AA'))\},$$

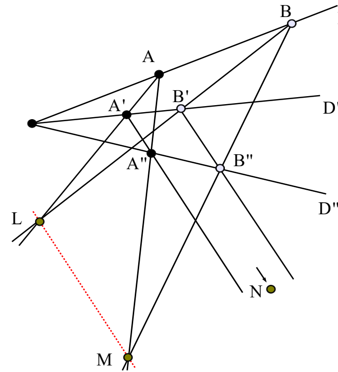
on dénombre 6 cas de figures différentes dans Desargues :

▷ Si D, D' et D'' se coupent dans le plan affine (où l'on dessine),
- Cas n°1 : Tous les couples de $\mathcal{T}$ sont formés de droites sécantes,
- Cas n°2 : Seulement deux couples de $\mathcal{T}$ sont formés de droites sécantes,
- Cas n°3 : Tous les couples de $\mathcal{T}$ sont formés de droites parallèles.

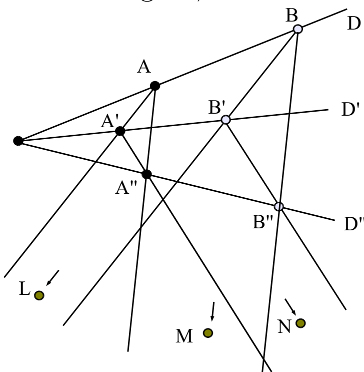
▷ Si D, D' et D'' se coupent à l'infini,
- Cas n°4 : Tous les couples de $\mathcal{T}$ sont formés de droites sécantes,
- Cas n°5 : Seulement deux couples de $\mathcal{T}$ sont formés de droites sécantes,
- Cas n°6 : Tous les couples de $\mathcal{T}$ sont formés de droites parallèles.



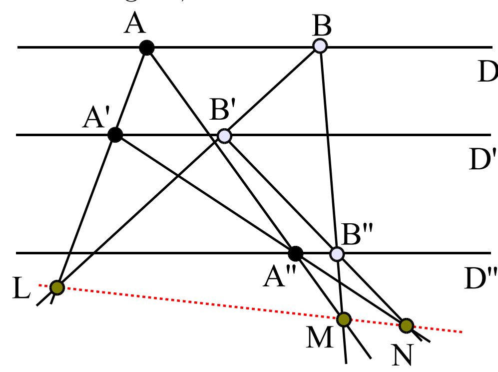
Desargues, Cas n°1



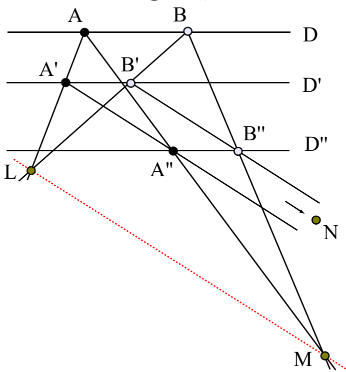
Desargues, Cas n°2



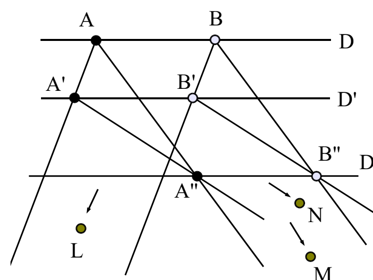
Desargues, Cas n°3



Desargues, Cas n°4



Desargues, Cas n°5



Desargues, Cas n°6

Les figures obtenues aux Cas n°3 et n°6 sont celles d'un autre énoncé dû à Desargues concernant les triangles homothétiques (deux triangles sont dits homothétiques s'ils se déduisent l'un de l'autre par une homothétie-translation). Cet autre "Théorème de Desargues" nous donne une condition nécessaire et suffisante pour que deux triangles soient homothétiques. La démonstration étant facile, elle peut être retenue. Il s'agit du Théorème 3.14 :

Théorème 3.14 *Dans un espace affine de dimension quelconque, deux triangles ABC et $A'B'C'$ sont homothétiques si et seulement si leurs côtés sont deux à deux parallèles.*

Preuve : La condition est évidemment nécessaire puisqu'une homothétie-translation transforme une droite en une droite parallèle. Montrons qu'elle est suffisante. Supposons que les trois couples $((AB),(A'B'))$, $((BC),(B'C'))$ et $((CA),(C'A'))$ soient formés de droites parallèles entre elles.

Si $A = A'$, ou $B = B'$, ou $C = C'$, il est facile de vérifier qu'une homothétie transforme les sommets de ABC en ceux de $A'B'C'$ (si $A = A'$, l'homothétie de centre A et de rapport $\overline{AB'}/\overline{AB}$ transforme B en B', mais aussi C en C').

Supposons maintenant que les droites (AA'), (BB') et (CC') sont bien définies. Puisque (AB) est parallèle à $(A'B')$, les points A, B, A', B' sont coplanaires, et les droites (AA') et (BB') seront parallèles ou sécantes. On envisage deux cas :

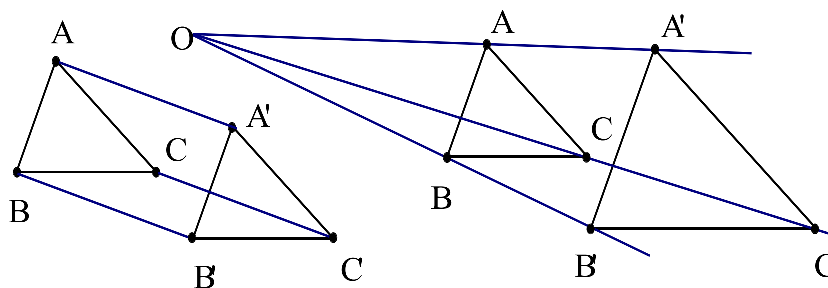


FIG. 3.11 – CNS pour que 2 triangles soient homothétiques

• Si les trois droites (AA'), (BB') et (CC') sont parallèles, les quadrilatères $AA'B'B$ et $BB'C'C$ sont des parallélogrammes, et la translation t de vecteur $\overrightarrow{AA'}$ amène A, B, C resp. sur A', B', C'.

• Sinon, deux des trois droites sont sécantes en un point O. Supposons par exemple que (AA') coupe (BB') en O. L'homothétie h de centre O et de

rapport $\overline{OA'}/\overline{OA}$ amène A sur A', mais aussi B sur B'. h transforme la droite (AC) en une droite parallèle à (AC) passant par A', c'est-à-dire en $(A'C')$. De même h transforme (BC) en $(B'C')$. Puisque h est bijective,

$$\begin{aligned} h(\{C\}) &= h((AC)\cap(BC)) \\ &= h((AC))\cap h((BC)) = \left(A'C'\right)\cap\left(B'C'\right) = \left\{C'\right\}, \end{aligned}$$

donc $h(C) = C'$. ∎

3.9 Dualité

3.9.1 Le principe

Soit $E = \mathbb{P}(\overrightarrow{E})$ un espace projectif de dimension 2 associé à l'espace vectoriel $\overrightarrow{E}$. Soit $p : \overrightarrow{E} \rightarrow E$ la projection canonique de $\overrightarrow{E}$ sur E. On sait que l'ensemble $\overrightarrow{E}^*$ des formes linéaires définies sur $\overrightarrow{E}$ forme un espace vectoriel de même dimension que $\overrightarrow{E}$.
On considère l'espace projectif $E^* = \mathbb{P}(\overrightarrow{E}^*)$ associé à $\overrightarrow{E}^*$, et l'on note encore (abusivement) p la projection canonique $p : \overrightarrow{E}^* \rightarrow E^*$ (FIG. 3.12).

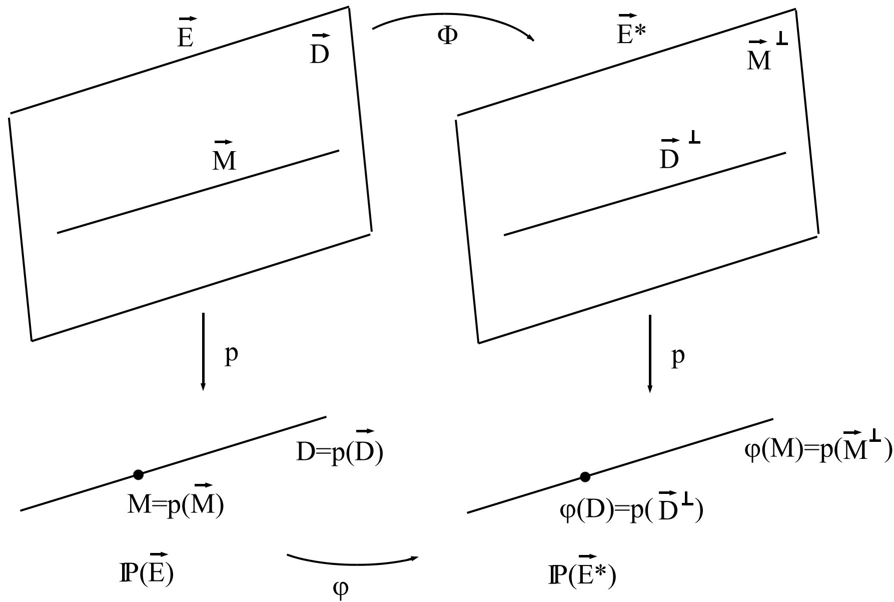


FIG. 3.12 – D'un espace à l'autre

On définit une fonction Φ qui à des droites et des plans de $\overrightarrow{E}$ associe des droites et des plans de $\overrightarrow{E}^*$, de la façon suivante :

a) A une droite $\overrightarrow{M}$ de $\overrightarrow{E}$ on associe le plan $\overrightarrow{M}^\perp$ de $\overrightarrow{E}^*$ constitué par toutes les formes linéaires sur $\overrightarrow{E}$ qui s'annulent sur $\overrightarrow{M}$ (l'orthogonalité est à considérer ici dans le cadre de la dualité),

b) A un plan $\overrightarrow{D}$ de $\overrightarrow{E}$ on associe la droite $\overrightarrow{D}^\perp$ de $\overrightarrow{E}^*$ constituée par toutes les formes linéaires sur $\overrightarrow{E}$ qui s'annulent sur $\overrightarrow{D}$.

Cette correspondance nous permet de définir une application φ qui à des points et des droites de E associe des droites et des points de E^*, de la façon suivante :

a) A un point $M = p(\overrightarrow{M})$ de E on associe la droite $\varphi(M) = p(\overrightarrow{M}^\perp)$ de E^*, projection de $\Phi(\overrightarrow{M})$ sur E^*,

b) A une droite $D = p(\overrightarrow{D})$ de E on associe le point $\varphi(D) = p(\overrightarrow{D}^\perp)$ de E^*, projection de $\Phi(\overrightarrow{D})$ sur E^*.

On vérifie alors :

Théorème 3.15
(P1) *Par φ, l'image d'un point est une droite, et l'image d'une droite est un point.*
(P2) *Si M appartient à la droite D, alors $\varphi(M)$ contient $\varphi(D)$.*

La propriété **(P2)** s'écrit

$$M \in D \ \Rightarrow\ \varphi(D) \in \varphi(M)$$

et montre que si M et N appartiennent à la droite D, alors $\varphi(D)$ est un point situé à l'intersection des droites $\varphi(M)$ et $\varphi(N)$.

Par exemple, la FIG. 3.13 montre que la correspondance φ fait passer de la configuration de Ceva et celle de Ménélaüs, et réciproquement. Les deux configurations sont donc duales l'une de l'autre.

De façon générale, dès qu'il s'agit de problèmes d'alignement et de concours, l'utilisation de la dualité, qui permet de passer d'une figure (formée de points et de droites) à une autre figure (en échangeant les rôles des points et des droites) permet de déduire mécaniquement de nouveaux résultats qui peuvent être très intéressants. Nous allons voir cette méthode en oeuvre sur deux exemples.

3.9.2 Le dual du Théorème de Pappus

La correspondance φ utilisée sur la configuration du Théorème de Pappus nous permet d'obtenir un nouveau Théorème qui est loin d'être une trivialité :

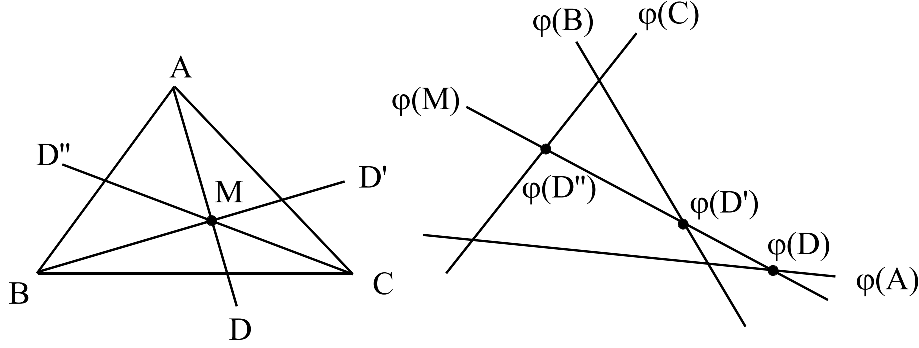


FIG. 3.13 – Dualité : Céva et Ménélaüs

Théorème 3.16 *(Dual du Théorème de Pappus) On considère trois droites a, b, c sécantes en d, ainsi que trois autre droites a', b', c' sécantes en d'. Alors les trois droites joignant $a \cap b'$ et $a' \cap b$, $b \cap c'$ et $b' \cap c$, puis $c \cap a'$ et $c' \cap a$, sont concourantes.*

Preuve : La FIG. 3.14 nous montre le passage de la configuration de Pappus, dans le dessin du haut, vers la nouvelle configuration dans le dessin du bas, via la correspondance φ.

On traduit les hypothèses. Dans Pappus, on dispose de deux droites D et D' contenant chacune trois points : A, B, C pour D, et A', B', C' pour D'. D'après le Théorème 3.15, $\varphi(D)$ sera un point appartenant à chacune des droites $\varphi(A)$, $\varphi(B)$, $\varphi(C)$. Et de même, $\varphi(D')$ sera un point appartenant à chacune des droites $\varphi(A')$, $\varphi(B')$, $\varphi(C')$.
Le point L appartient à (AB') et à $(A'B)$. On a

$$L \in \left(AB'\right) \;\Rightarrow\; \varphi\left(\left(AB'\right)\right) \in \varphi(L),$$

et l'on remarque que $\varphi((AB'))$ est un point situé à l'intersection des droites $\varphi(A)$ et $\varphi(B')$. Ainsi la droite $\varphi(L)$ passe par l'intersection des droites $\varphi(A)$ et $\varphi(B')$.
En recommençant de la même façon, mais en partant de l'implication

$$L \in \left(A'B\right) \;\Rightarrow\; \varphi\left(\left(A'B\right)\right) \in \varphi(L),$$

on montre aussi que $\varphi(L)$ passe par l'intersection des droites $\varphi(A')$ et $\varphi(B)$. En conclusion $\varphi(L)$ est la droite joignant $\varphi(A) \cap \varphi(B')$ et $\varphi(A') \cap \varphi(B)$. De la même manière, on constate que :

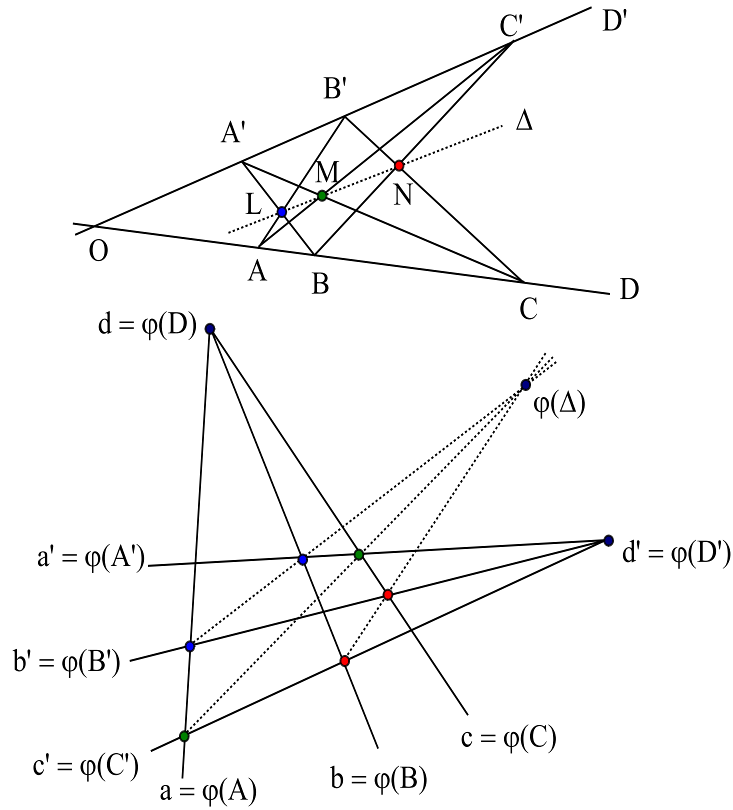


FIG. 3.14 – Le Théorème de Pappus et le résultat dual

- $\varphi(M)$ est la droite joignant $\varphi(A) \cap \varphi(C')$ et $\varphi(A') \cap \varphi(C)$,

- $\varphi(N)$ est la droite joignant $\varphi(B) \cap \varphi(C')$ et $\varphi(B') \cap \varphi(C)$.

La situation est bien celle décrite dans le Théorème (et décrite sur la FIG. 3.14).

Il est temps de traduire la conclusion du Théorème de Pappus selon laquelle il existe une droite Δ contenant les points L, M et N. La traduction donne cela : il existe un point $\varphi(\Delta)$ appartenant à l'intersection $\varphi(L) \cap \varphi(M) \cap \varphi(N)$.

Les trois droites $\varphi(L)$, $\varphi(M)$ et $\varphi(N)$ sont donc bien concourantes ! ∎

3.9.3 Réciproque du Théorème de Desargues

Quel est l'énoncé dual du Théorème de Desargues ? Le Théorème de Desargues (Th. 3.13) s'écrit :

$$\begin{cases} O \in (AB) \\ O \in (A'B') \\ O \in (A''B'') \end{cases} \Rightarrow \exists \text{ droite } \Delta \quad \begin{cases} (AA') \cap (BB') \subset \Delta \\ (AA'') \cap (BB'') \subset \Delta \\ (A'A'') \cap (B'B'') \subset \Delta \end{cases} \qquad (*)$$

(voir FIG. 3.15). L'énoncé dual est donc :

$$\begin{cases} \varphi(AB) \in \varphi(O) \\ \varphi(A'B') \in \varphi(O) \\ \varphi(A''B'') \in \varphi(O) \end{cases} \Rightarrow \exists \text{point } \varphi(\Delta) \quad \begin{cases} \varphi(\Delta) \in \varphi((AA') \cap (BB')) \\ \varphi(\Delta) \in \varphi((AA'') \cap (BB'')) \\ \varphi(\Delta) \in \varphi((A'A'') \cap (B'B'')) . \end{cases} \qquad (\maltese)$$

[Remarque au sujet des notations : pour simplifier, je noterai $\varphi(AA')$ au lieu de $\varphi((AA'))$ le point image de la droite (AA') par φ, et $(\varphi(AA')\varphi(BB'))$ la droite passant par les points $\varphi(AA')$ et $\varphi(BB')$, et ainsi de suite... Je m'autoriserai aussi à noter indifféremment M ou $\{M\}$ quand cela m'arrange, pour donner un sens aux écritures du style $\varphi((AA') \cap (BB'))$.]

La FIG. 3.15 nous permet de comprendre l'analogie entre les assertions $(*)$ et $(\maltese)$. On constate par exemple que

$$\begin{cases} \varphi((AA') \cap (BB')) = (\varphi(AA')\varphi(BB')) \\ \varphi((AA'') \cap (BB'')) = (\varphi(AA'')\varphi(BB'')) \\ \varphi((A'A'') \cap (B'B'')) = (\varphi(A'A'')\varphi(B'B'')) , \end{cases} \qquad (1)$$

ce qui se vérifie en rappelant que la droite $\varphi((AA') \cap (BB'))$ contient $\varphi(AA')$ et $\varphi(BB')$, et ainsi de suite. La FIG. 3.15 suggère aussi que :

$$\begin{cases} \{\varphi(AB)\} = (\varphi(AA')\varphi(AA'')) \cap (\varphi(BB')\varphi(BB'')) \\ \{\varphi(A'B')\} = (\varphi(AA')\varphi(A'A'')) \cap (\varphi(BB')\varphi(B'B'')) \\ \{\varphi(A''B'')\} = (\varphi(AA'')\varphi(A'A'')) \cap (\varphi(BB'')\varphi(B'B'')) . \end{cases} \qquad (2)$$

Pour justifier par exemple la première égalité de (2), on remarque que $\varphi(A)$ est une droite qui passe par les points $\varphi(AA')$ et $\varphi(AA'')$, et que par conséquent $\varphi(A) = (\varphi(AA')\varphi(AA''))$. De la même façon $\varphi(B) = (\varphi(BB')\varphi(BB''))$ et :

$$\big(\varphi(AA')\varphi(AA'')\big) \cap \big(\varphi(BB')\varphi(BB'')\big) = \varphi(A) \cap \varphi(B) = \{\varphi(AB)\}.$$

En utilisant (1) et (2), l'implication $(\maltese)$ s'écrit :

$$\begin{cases} (\varphi(AA')\varphi(AA'')) \cap (\varphi(BB')\varphi(BB'')) \subset \varphi(O) \\ (\varphi(AA')\varphi(A'A'')) \cap (\varphi(BB')\varphi(B'B'')) \subset \varphi(O) \\ (\varphi(AA'')\varphi(A'A'')) \cap (\varphi(BB'')\varphi(B'B'')) \subset \varphi(O) \end{cases}$$

$$\Rightarrow \exists \text{ point } \varphi(\Delta) \quad \begin{cases} \varphi(\Delta) \in (\varphi(AA')\varphi(BB')) \\ \varphi(\Delta) \in (\varphi(AA'')\varphi(BB'')) \\ \varphi(\Delta) \in (\varphi(A'A'')\varphi(B'B'')) . \end{cases}$$

On reconnaît l'implication réciproque de l'implication $(*)$. En conclusion :

Théorème 3.17 *L'énoncé dual du Théorème de Desargues n'est autre que la réciproque du théorème.*

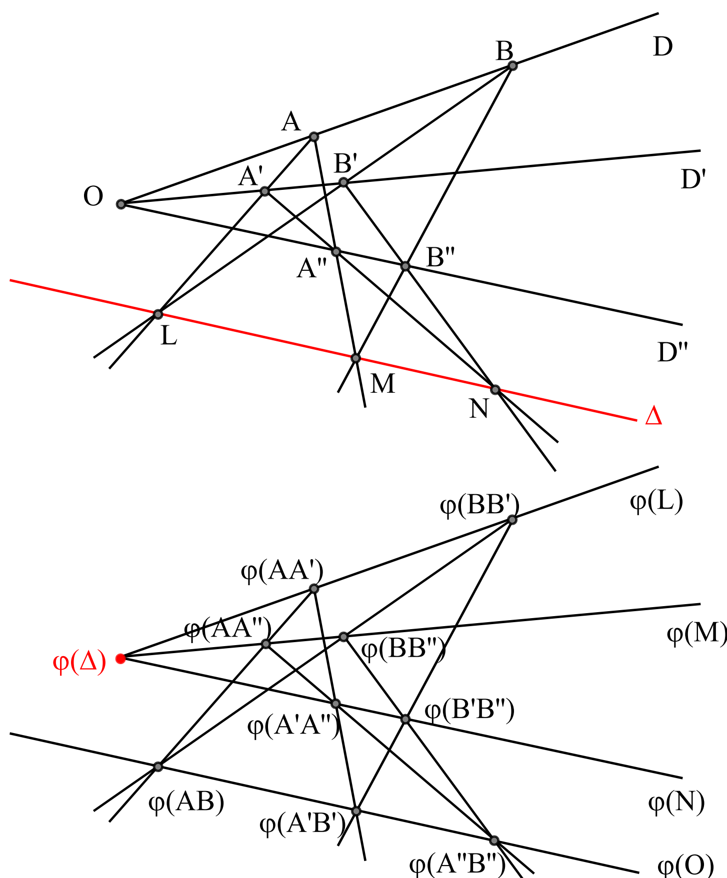


FIG. 3.15 – Dual du Théorème de Desargues

On dit que le Théorème de Desargues est auto-dual pour signifier que sa réciproque est exactement l'énoncé dual. Et finalement, on a bien démontré cette réciproque en utilisant la dualité. Enonçons-la :

Théorème 3.18 ***(Réciproque du Théorème de Desargues)*** *Dans le plan projectif, on considère trois droites distinctes D, D', D'', et six points distincts placés sur chacune de ces droites : A, B sur D ; A', B' sur D' ; et A'', B'' sur D''. On note L, M et N les points d'intersection des couples de droites*

$((AA'),(BB'))$, $((AA''),(BB''))$ et $(A'A''),(B'B'')$. Si les points L, M, N sont alignés, alors les droites D, D', D'' sont concourantes.

3.10 Homographies

3.10.1 Définitions

Toute application linéaire injective u d'un espace vectoriel E vers un espace vectoriel F se factorise en une application $P(u)$ entre les espaces projectifs $\mathbb{P}(E)$ et $\mathbb{P}(F)$. En effet, si $\dot{x} = \dot{y}$ il existe $\lambda \in K^*$ tel que $y = \lambda x$, donc $u(y) = \lambda u(x)$. u étant injective, aucun des vecteurs $u(x)$ ou $u(y)$ n'est nul, et l'on peut écrire

$$\dot{\overline{u(y)}} = \dot{\overline{u(x)}}.$$

On peut donc poser

$$P(u)\left(\dot{x}\right) = \dot{\overline{u(x)}}.$$

Le diagramme suivant est commutatif :

$$\begin{array}{ccc} E & \xrightarrow{u} & F \\ {\scriptstyle p}\downarrow & & \downarrow{\scriptstyle p} \\ \mathbb{P}(E) & \underset{P(u)}{\longrightarrow} & \mathbb{P}(F) \end{array}$$

Définition 3.9 *L'application $P(u)$ est **l'application projective associée à** u. Une application projective bijective est une **homographie**.*

3.10.2 Premières propriétés

Théorème 3.19 *Une application projective $P(u)$ est toujours injective.*

Preuve : En effet

$$\begin{aligned} P(u)\left(\dot{x}\right) = P(u)\left(\dot{y}\right) & \Rightarrow \quad \dot{\overline{u(y)}} = \dot{\overline{u(x)}} \\ & \Rightarrow \quad \exists \lambda \in K^* \quad u(y) = \lambda u(x) \\ & \Rightarrow \quad \exists \lambda \in K^* \quad y = \lambda x \quad \text{(car } u \text{ injective)} \\ & \Rightarrow \quad \dot{x} = \dot{y}. \blacksquare \end{aligned}$$

Théorème 3.20 *L'application linéaire injective u est bijective si et seulement si $P(u)$ est bijective.*

Preuve : Comme u et $P(u)$ sont toujours injectives, il reste seulement à vérifier que u est surjective si et seulement si $P(u)$ l'est.
Si u est surjective, et si $\dot{y} \in \mathbb{P}(F)$, il existe $x \in E$ tel que $u(x) = y$. Alors

$$P(u)\left(\dot{x}\right) = \overset{\cdot}{\overline{u(x)}} = \dot{y},$$

et $\dot{y}$ possède au moins un antécédent dans $\mathbb{P}(E)$.
Réciproquement, si $P(u)$ est surjective et si y est un vecteur quelconque de F, il existe $\dot{x} \in \mathbb{P}(E)$ tel que

$$P(u)\left(\dot{x}\right) = \dot{y},$$

d'où $\overset{\cdot}{\overline{u(x)}} = \dot{y}$ et l'existence d'un scalaire non nul λ tel que $y = \lambda u(x)$. Par suite $y = u(\lambda x)$ possède bien au moins un antécédent par u. ∎

Théorème 3.21 *La composée de deux applications projectives est encore une application projective. Plus précisément, si $u : E \to F$ et $v : F \to G$ sont deux applications linéaires injectives, alors $P(v \circ u) = P(v) \circ P(u)$, et bien sûr $P(Id) = Id$.*

Preuve : Remarquons d'abord que l'application $P(v \circ u)$ est bien définie puisque $v \circ u$ est injective comme composée de deux injections.
Pour tout $\dot{x} \in \mathbb{P}(E)$,

$$P(v \circ u)(\dot{x}) = \overset{\cdot}{\overline{v(u(x))}} = P(v)\left(\overset{\cdot}{\overline{u(x)}}\right) = P(v)\left[P(u)\left(\dot{x}\right)\right] = P(v \circ u)\left(\dot{x}\right),$$

donc $P(v \circ u) = P(v) \circ P(u)$. Par ailleurs $P(Id)(\dot{x}) = \overset{\cdot}{\overline{Id(x)}} = \dot{x}$ pour tout $\dot{x} \in \mathbb{P}(E)$, donc $P(Id) = Id$. ∎

Théorème 3.22 *La composée de deux homographies est une homographie. L'ensemble des homographies de $\mathbb{P}(E)$ dans $\mathbb{P}(E)$ est un groupe pour la composition des applications.*

Preuve : L'ensemble des homographies de $\mathbb{P}(E)$ dans $\mathbb{P}(E)$ n'est pas vide (il contient $P(Id) = Id$), est stable par composition d'après le Théorème 3.21, et l'application réciproque d'une homographie $P(u)$ est encore une homographie : c'est $P(u^{-1})$ puisque

$$P(u^{-1}) \circ P(u) = P(u) \circ P(u^{-1}) = P(Id) = Id. \blacksquare$$

Définition 3.10 *Le groupe des homographies de $\mathbb{P}(E)$ dans $\mathbb{P}(E)$ est noté* $\mathrm{GP}(E)$ *et appelé* ***groupe projectif*** *de l'espace vectoriel E.*

Théorème 3.23 *Si u et v sont deux applications linéaires injectives de E vers F, alors*

$$P(u) = P(v) \quad \Leftrightarrow \quad \exists k \in K^* \quad v = ku.$$

En particulier $P(u) = Id$ si, et seulement si, u est une homothétie vectorielle de rapport non nul, et $\mathrm{GP}(E) \simeq \mathrm{GL}(E)/K^*$.

Preuve : La condition est suffisante. Montrons sa nécessité.
Dire que $P(u) = P(v)$ revient à dire que pour tout $x \in E$, les vecteurs $u(x)$ et $v(x)$ sont colinéaires. Il existe donc $k_x \in K$ tel que

$$\forall x \in E \quad v(x) = k_x.u(x).$$

On vérifie que k_x est indépendant de x.

• Si x et y sont linéairement indépendants, $u(x)$ et $u(y)$ le seront aussi et

$$\begin{aligned} v(x+y) = v(x) + v(y) \quad &\Rightarrow \quad k_{x+y}.u(x+y) = k_x.u(x) + k_y.u(y) \\ &\Rightarrow \quad k_{x+y}.u(x) + k_{x+y}.u(y) = k_x.u(x) + k_y.u(y) \\ &\Rightarrow \quad k_{x+y} = k_y = k_x. \end{aligned}$$

• Si x et y sont liés et x non nul, et si $\dim E \geq 2$, on peut toujours passer par l'intermédiaire d'un vecteur z n'appartenant pas à la droite $\mathrm{Vect}(x)$ et appliquer deux fois le résultat précédent pour obtenir $k_x = k_z = k_y$. Si $\dim E = 1$, on note $y = \lambda x$ et l'on conclut en écrivant :

$$\begin{aligned} v(y) = \lambda v(x) \quad &\Rightarrow \quad k_y.u(y) = \lambda k_x.u(x) \\ &\Rightarrow \quad \lambda k_y.u(x) = \lambda k_x.u(x) \\ &\Rightarrow \quad k_y = k_x. \end{aligned}$$

Pour terminer, on montre que $\mathrm{GP}(E) \simeq \mathrm{GL}(E)/K^*$ en décomposant canoniquement le morphisme de groupes $\psi : (\mathrm{GL}(E), \circ) \rightarrow (\mathrm{GP}(E), \circ)$ qui à u fait correspondre $P(u)$. Dans la notation $\mathrm{GL}(E)/K^*$, il faut être conscient que K^* désigne le noyau de ψ, c'est-à-dire le groupe des homothéties de E (et que la notation est un tantinet cavalière !). ∎

Remarque : On aurait pu court-circuiter une partie de la démonstration précédente en rappelant le lemme d'algèbre linéaire bien connu suivant lequel tout endomorphisme qui laisse stable toutes les droites vectorielles est une homothétie ([1], Th. 33).

Théorème 3.24 *Une application projective transforme une droite en une droite (et en particulier, elle conserve les alignements de points, c'est-à-dire transforme trois points alignés en trois points alignés).*

Preuve : Considérons une application projective $P(u) : \mathbb{P}(E) \longrightarrow \mathbb{P}(F)$, et une droite $(\dot{x}\dot{y})$ de $\mathbb{P}(E)$. Ici $(\dot{x}\dot{y})$ désigne la droite $p(\Pi)$ où $\Pi = \mathrm{Vect}(x, y)$ est le plan vectoriel de E engendré par x et y. Si $\dot{z} \in (\dot{x}\dot{y})$, il existe deux scalaires λ et μ tels que $z = \lambda x + \mu y$, donc

$$P(u)(\dot{z}) = \dot{\overline{u(z)}} = \dot{\overline{\lambda u(x) + \mu u(y)}}.$$

Cela montre que $u(z) \in \mathrm{Vect}(u(x), u(y))$, autrement dit que les points $\dot{\overline{u(z)}}$, $\dot{\overline{u(x)}}$, et $\dot{\overline{u(y)}}$ sont alignés. Le point $P(u)(\dot{z})$ appartient par conséquent à la droite $(\dot{\overline{u(x)}}\ \dot{\overline{u(y)}})$. On a montré l'inclusion

$$P(u)\left((\dot{x}\dot{y})\right) \subset (\dot{\overline{u(x)}}\,\dot{\overline{u(y)}}).$$

Réciproquement, si $\dot{z}'$ appartient à la droite $(\dot{\overline{u(x)}}\ \dot{\overline{u(y)}})$, z' est un vecteur de $\mathrm{Vect}(u(x), u(y))$ et il existe λ et μ tels que $z' = \lambda u(x) + \mu u(y) = u(\lambda x + \mu y)$. Par suite

$$P(u)(\dot{\overline{\lambda x + \mu y}}) = \dot{\overline{u(\lambda x + \mu y)}} = \dot{z}'$$

où $\dot{\overline{\lambda x + \mu y}}$ est un point de la droite $(\dot{x}\dot{y})$, donc $\dot{z}' \in P(u)\left((\dot{x}\dot{y})\right)$, et l'inclusion réciproque est démontrée. ■

Remarque : On démontre que $\mathrm{GP}\left(\mathbb{R}^{n+1}\right)$ est égal au groupe des collinéations, c'est-à-dire au groupe des bijections qui conservent l'alignement. Ce n'est pas le cas du groupe projectif $\mathrm{GP}(E)$ en général.

3.10.3 Conservation des repères projectifs

Théorème 3.25 *Soient* $(P_0, P_1, ..., P_{n+1})$ *et* $(Q_0, Q_1, ..., Q_{n+1})$ *des repères projectifs respectifs de* $\mathbb{P}(E)$ *et de* $\mathbb{P}(F)$*. Il existe une et une seule homographie de* $\mathbb{P}(E)$ *dans* $\mathbb{P}(F)$ *qui envoie* P_i *sur* Q_i *pour tout* i*.*

Preuve : Par hypothèse (Définition 3.5), il existe une base $e = (e_0, ..., e_n)$ de E et une base $f = (f_0, ..., f_n)$ de F telles que :

- $\forall i \in \{0, 1, ..., n\} \quad P_i = p(e_i)$ et $Q_i = p(f_i)$;
- $P_{n+1} = p\left(\sum_{i=0}^n e_i\right)$ et $Q_{n+1} = p\left(\sum_{i=0}^n f_i\right)$.

S'il existe une homographie $\alpha = P(u) : \mathbb{P}(E) \to \mathbb{P}(F)$ satisfaisant aux conditions du Théorème, alors $u(e_i)$ et f_i sont colinéaires quels que soient les indices i, et il existe des scalaires λ_i tels que $u(e_i) = \lambda_i f_i$. Mais il existe aussi un scalaire λ tel que $u\left(\sum_{i=0}^n e_i\right) = \lambda \sum_{i=0}^n f_i$, ce qui s'écrit

$$\sum_{i=0}^n \lambda_i f_i = \lambda \sum_{i=0}^n f_i.$$

Par suite $\lambda_0 = ... = \lambda_n = \lambda$. L'application linéaire u, qui vérifie $u(e_i) = \lambda f_i$ pour tout i, est ainsi parfaitement déterminée sur chacun des vecteurs de la base e, donc unique. L'unicité de α est démontrée.
Pour vérifier qu'il existe au moins une homographie α solution, on considère l'application linéaire u de E dans F (obtenue précédemment) qui transforme la base e en la base f, et l'on pose $\alpha = P(u)$. Bien sûr

$$\forall i \in \{0, ..., n\} \quad \alpha(P_i) = P(u)(P_i) = p(u(e_i)) = p(f_i) = Q_i,$$

mais aussi :

$$\alpha(P_{n+1}) = P(u)(P_{n+1}) = p\left(u\left(\sum_{i=0}^{n} e_i\right)\right) = p\left(\sum_{i=0}^{n} f_i\right) = Q_{n+1}. \blacksquare$$

3.10.4 Lien avec les fonctions homographiques

▶ **Cas de la droite projective** $\mathbb{P}(\mathbb{R}^2)$

Travaillons dans $\mathbb{P}(\mathbb{R}^2)$ rapporté à un repère projectif dans lequel tout point de $\mathbb{P}(\mathbb{R}^2)$ admet des coordonnées homogènes $(x : y)$. Conservons les notations de la Section 3.5 en les simplifiant un peu : A désigne le complémentaire de l'hyperplan H de $\mathbb{P}(\mathbb{R}^2)$ d'équation $y = 0$, donc

$$A = \left\{M(x : y) \in \mathbb{P}(\mathbb{R}^2) \,/\, y \neq 0\right\}.$$

On sait que A est un espace affine de dimension 1, donc isomorphe à $\mathbb{R}$. La bijection f qui permet d'identifier A à $\mathbb{R}$ est

$$\begin{array}{rccc} f: & K & \to & A = \mathbb{P}(\mathbb{R}^2)\backslash H \\ & x & \mapsto & (x:1). \end{array}$$

Une homographie de $\mathbb{P}(\mathbb{R}^2)$ s'écrit $P(u)$ où $u \in \mathrm{GL}\left(\mathbb{R}^2\right)$ est défini par une matrice de déterminant non nul dans la base canonique de $\mathbb{R}^2$. Autrement dit, l'endomorphisme u est définie analytiquement en écrivant

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

où $ad - bc \neq 0$. On a

$$P(u)(x : y) = \overline{(x', y')}^{\,\cdot} = \left(x' : y'\right).$$

La question est de savoir quelle est la restriction $h = P(u)|_A$ de $P(u)$ à A, et nous ne voulons pas seulement restreindre $P(u)$ au départ, mais aussi à l'arrivée pour nous intéresser à l'application

$$h = P(u)|_A : A \to A.$$

La question est donc de savoir si, étant donné $(x : y)$ tel que $y \neq 0$, le réel $y' = cx + dy$ est nul ou pas. On a

$$y' = 0 \quad \Leftrightarrow \quad cx + dy = 0 \quad \Leftrightarrow \quad \frac{x}{y} = -\frac{d}{c}.$$

Par suite, si $c \neq 0$, pour tout $t = x/y$ différent de $-d/c$ (où $y \neq 0$), on peut écrire

$$P(u)(t : 1) = P(u)\left(\frac{x}{y} : 1\right) = \left(\frac{x'}{y'} : 1\right) = \left(\frac{ax + by}{cx + dy} : 1\right) = \left(\frac{at + b}{ct + d} : 1\right).$$

Compte tenu de l'identification de A à $\mathbb{R}$ (détaillée à la Section 3.5 et rappelée ci-dessus), $h = P(u)|_A : A \to A$ apparaît comme une fonction de $\mathbb{R}$ dans $\mathbb{R}$ qui à t (différent de $-d/c$) associe

$$h(t) = \frac{at + b}{ct + d}.$$

Une telle fonction

$$\begin{array}{rccc} h: & \mathbb{R} & \to & \mathbb{R} \\ & t & \mapsto & \dfrac{at + b}{ct + d} \end{array}$$

est appelée *fonction homographique*.

Si $c = 0$, alors d ne peut pas être nul (car $ad - bc \neq 0$), donc $ct + d = d$ n'est jamais nul, et h est définie sur tout $\mathbb{R}$ (la fonction h devient une application de $\mathbb{R}$ dans $\mathbb{R}$). C'est l'application affine :

$$\begin{array}{rccc} h: & \mathbb{R} & \to & \mathbb{R} \\ & t & \mapsto & \dfrac{at + b}{d}. \end{array}$$

Remarque : Comme $P(v \circ u) = P(v) \circ P(u)$, on peut estimer que la composée $g \circ h$ de deux fonctions homographiques

$$t \xrightarrow{h} \frac{at + b}{ct + d} \quad \text{et} \quad t \xrightarrow{g} \frac{a't + b'}{c't + d'}$$

sera une fonction homographique restriction de $P(v \circ u)$ à la carte affine (lorsqu'on met de côté tous les problèmes de définition de ces fonctions). Eh bien, c'est le cas puisque l'on peut vérifier que

$$g \circ h : t \to \frac{a''t + b''}{c''t + d''} \quad \text{avec} \quad \begin{pmatrix} a'' & b'' \\ c'' & d'' \end{pmatrix} = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

▶ **Cas général**

Ce qui précède se généralise facilement au cas général où $P(u)$ est une homographie de $\mathbb{P}(\mathbb{R}^{n+1})$. Dans ce cas $P(u)(x_0 : x_1 : ... : x_n) = (x'_0 : x'_1 : ... : x'_n)$ où

$$\begin{pmatrix} x'_0 \\ \vdots \\ x'_n \end{pmatrix} = \begin{pmatrix} a_{00} & ... & a_{0n} \\ \vdots & & \vdots \\ a_{n0} & ... & a_{nn} \end{pmatrix} \begin{pmatrix} x_0 \\ \vdots \\ x_n \end{pmatrix}.$$

En supposant $x_n \neq 0$ et $x'_n \neq 0$, et en posant $t_i = x_i/x_n$ $(0 \leq i \leq n-1)$, on obtient

$$\begin{aligned} P(u)(t_0 : t_1 : ... : t_{n-1} : 1) &= \left(\frac{x'_0}{x'_n} : \frac{x'_1}{x'_n} : ... : \frac{x'_{n-1}}{x'_n} : 1\right) \\ &= (y_0 : y_1 : ... : y_{n-1} : 1) \end{aligned}$$

avec

$$y_i = \frac{x'_i}{x'_n} = \frac{a_{i0}x_0 + ... + a_{in}x_n}{a_{n0}x_0 + ... + a_{nn}x_n} = \frac{a_{i0}t_0 + ... + a_{i,n-1}t_{n-1} + a_{in}}{a_{n0}t_0 + ... + a_{i,n-1}t_{n-1} + a_{nn}}.$$

Finalement, les restrictions des homographies aux cartes affines de l'espace projectif sont définies analytiquement avec des fonctions coordonnées qui sont des quotients de formes affines par une même forme affine.

3.10.5 Conservation du birapport

Si les applications affines conservent les rapports de mesures algébriques des bipoints de points alignés (on dit plus rapidement qu'une application affine conserve les rapports), on montre que les homographies conservent les rapports de rapports, c'est-à-dire les *birapports*. Ces birapports représentent donc des invariants de la géométrie projective.

Considérons une droite projective $\mathcal{D}$ d'un espace projectif $\mathbb{P}\left(K^{n+1}\right)$, et quatre points distincts M_1, M_2, M_3, M_4 sur cette droite. Considérons un repère projectif $\mathcal{R}$ de $\mathcal{D}$, notons $(x_i : y_i)$ les coordonnées homogènes de M_i dans $\mathcal{R}$, et posons

$$[M_1, M_2, M_3, M_4]_{\mathcal{R}} = \frac{\begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix}}{\begin{vmatrix} x_1 & x_4 \\ y_1 & y_4 \end{vmatrix}} : \frac{\begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix}}{\begin{vmatrix} x_2 & x_4 \\ y_2 & y_4 \end{vmatrix}}.$$

Alors :

Théorème 3.26 *Le scalaire* $[M_1, M_2, M_3, M_4]_{\mathcal{R}}$ *est indépendant du choix du repère projectif* $\mathcal{R}$ *de la droite* $\mathcal{D}$.

Preuve : Si $\mathcal{R}'$ désigne un autre repère projectif de $\mathcal{D}$, et si $(x'_i : y'_i)$ désigne un système de coordonnées homogènes de M_i dans $\mathcal{R}'$, les formules de changement de bases dans le plan vectoriel Π associé à $\mathcal{D}$ (c'est-à-dire tel que $\mathcal{D} = p(\Pi)$) montre l'existence d'une matrice carrée inversible P telle que

$$\forall i, j \in \{1, 2, 3, 4\} \qquad \begin{pmatrix} x_i & x_j \\ y_i & y_j \end{pmatrix} = P \begin{pmatrix} x'_i & x'_j \\ y'_i & y'_j \end{pmatrix}.$$

Par suite

$$\forall i, j \in \{1, 2, 3, 4\} \qquad \begin{vmatrix} x_i & x_j \\ y_i & y_j \end{vmatrix} = (\det P) \times \begin{vmatrix} x'_i & x'_j \\ y'_i & y'_j \end{vmatrix}$$

et il suffit de remplacer et de simplifier par $\det P$ pour obtenir

$$[M_1, M_2, M_3, M_4]_{\mathcal{R}'} = [M_1, M_2, M_3, M_4]_{\mathcal{R}} . \blacksquare$$

Définition 3.11 *Le scalaire* $[M_1, M_2, M_3, M_4]_{\mathcal{R}}$*, noté* $[M_1, M_2, M_3, M_4]$*, est appelé* ***birapport*** *ou* ***rapport anharmonique*** *des points* M_1*,* M_2*,* M_3*,* M_4 *(alignés et distincts).*

L'invariance des birapports sous l'action des homographies se montre sans difficulté :

Théorème 3.27 *Une homographie conserve les birapports. Autrement dit, une homographie* $P(u) : \mathbb{P}(E) \to \mathbb{P}(F)$ *transforme quatre points* M_1*,* M_2*,* M_3*,* M_4 *alignés et distincts en quatre points* M'_1*,* M'_2*,* M'_3*,* M'_4 *alignés et distincts, et* $[M'_1, M'_2, M'_3, M'_4] = [M_1, M_2, M_3, M_4]$.

Preuve : Une homographie est bijective et conserve l'alignement, donc transforme quatre points alignés et distincts en quatre points alignés et distincts. Si M_1, M_2, M_3, M_4 sont alignés sur une droite $\mathcal{D}$ de $\mathbb{P}(E)$, les images M'_1, M'_2, M'_3, M'_4 de ces points par $P(u)$ seront alignées sur une droite $\mathcal{D}'$ de $\mathbb{P}(F)$.
Si $(x_i : y_i)$ désigne un système de coordonnées homogènes de M_i dans un repère $\mathcal{R}$ de $\mathcal{D}$, notons Π le plan vectoriel tel que $\mathcal{D} = p(\Pi)$ et $e = (e_1, e_2)$ une base de Π associée au repère $\mathcal{R}$.
L'isomorphisme u transforme la base e de Π en une base $u(e) = (u(e_1), u(e_2))$ du plan $\Pi' = u(\Pi)$, et l'on peut utiliser le repère $\mathcal{R}'$ de $\mathcal{D}'$ associé à cette base. On vérifie alors que les coordonnées homogènes de $M'_i = P(u)(M_i)$ dans $\mathcal{R}'$ sont encore $(x_i : y_i)$: en notant $M_i = p(\xi_i)$ où $\xi_i = x_i e_1 + y_i e_2 \in \Pi$, on obtient en effet

$$M'_i = P(u)(M_i) = p(u(\xi_i)) = p(x_i u(e_1) + y_i u(e_2)).$$

Finalement

$$\left[M_1', M_2', M_3', M_4'\right]_{\mathcal{R}'} = \frac{\begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix}}{\begin{vmatrix} x_1 & x_4 \\ y_1 & y_4 \end{vmatrix}} : \frac{\begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix}}{\begin{vmatrix} x_2 & x_4 \\ y_2 & y_4 \end{vmatrix}} = [M_1, M_2, M_3, M_4]_{\mathcal{R}}$$

et l'égalité des birapports est démontrée. ■

Nous avons défini le birapport de quatre points distincts et alignés dans le projectif, mais il y a un petit problème. En géométrie affine, nous avons déjà travaillé avec un birapport, qui était un rapport de deux rapports de mesures algébriques ([1], §. 16.1), et il est tout à fait naturel de nous poser la question de savoir si les deux définitions ne font qu'une là où elles ont toutes les deux un sens, c'est-à-dire lorsqu'on se restreint au complémentaire d'un hyperplan projectif muni de sa structure affine canonique.

Prenons $E = K^{n+1}$ et intéressons-nous à la trace d'une droite $\mathcal{D}$ de $\mathbb{P}\left(K^{n+1}\right)$ dans une carte affine $A = \mathbb{P}\left(K^{n+1}\right) \backslash H$ où H désigne un hyperplan projectif. On peut supposer que H est d'équation $x_n = 0$. Notons $\mathcal{D}_a = \mathcal{D} \cap A$, et considérons quatre point distincts M_1, M_2, M_3, M_4 sur $\mathcal{D}_a$. La FIG. 3.16 représente la droite affine $\mathcal{D}_a$ et le point à l'infini de $\mathcal{D}$, noté $\infty_{\mathcal{D}}$.

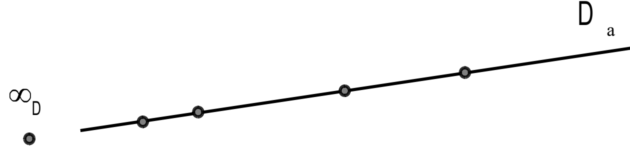


FIG. 3.16 – $\mathcal{D} = \mathcal{D}_a \cup \{\infty_{\mathcal{D}}\}$

Choisissons un repère projectif de $\mathbb{P}\left(K^{n+1}\right)$ dans lequel les coordonnées de $M_1 = p(e_0)$ et $M_2 = p(e_1)$ sont[5] :

$$\begin{cases} M_1 = (1:0:0:...:0:1) = p(e_0) \\ M_2 = (0:1:0:...:0:1) = p(e_1). \end{cases}$$

Les vecteurs e_0 et e_1 qui définissent M_1 et M_2 ne sont pas colinéaires puisque $M_1 \neq M_2$. Comme M_3 et M_4 appartiennent à la droite (M_1M_2), il existe des

[5] On identifie les points du projectif avec les coordonnées homogènes de ces points dans le repère considéré.

scalaires λ, μ, ν, τ tels que $\lambda+\mu=1$, $\nu+\tau=1$, et

$$\begin{cases} M_3 = p(\lambda e_0 + \mu e_1) = (\lambda : \mu : 0 : ... : 0 : 1) \\ M_4 = p(\nu e_1 + \tau e_1) \ = (\nu : \tau : 0 : ... : 0 : 1) \,. \end{cases}$$

Dans l'espace affine A (identifié à K^n) d'espace vectoriel associé $\overrightarrow{A}$ (identifié à K^n), les coordonnées affines des points M_i sont :

$$\begin{cases} M_1 = (1,0,0,..,0) \in K^n \\ M_2 = (0,1,0,...,0) \\ M_3 = (\lambda,\mu,0,...,0) \\ M_4 = (\nu,\tau,0,...,0) \,. \end{cases}$$

Par suite

$$\begin{cases} \overrightarrow{M_1M_3} = (\lambda-1,\mu,0,...,0) = \mu\,(-1,1,0,...,0) = \mu\overrightarrow{M_1M_2} \\ \overrightarrow{M_1M_4} = (\nu-1,\tau,0,...,0) = \tau\,(-1,1,0,...,0) = \tau\overrightarrow{M_1M_2} \end{cases}$$

donc

$$\frac{\overline{M_1M_3}}{\overline{M_1M_4}} = \frac{\mu}{\tau}. \quad (1)$$

De même

$$\begin{cases} \overrightarrow{M_2M_3} = (\lambda,\mu-1,0,...,0) = \lambda\,(1,-1,0,...,0) = \lambda\overrightarrow{M_2M_1} \\ \overrightarrow{M_2M_4} = (\nu,\tau-1,0,...,0) = \nu\,(-1,1,0,...,0) = \nu\overrightarrow{M_2M_1} \end{cases}$$

et

$$\frac{\overline{M_2M_3}}{\overline{M_2M_4}} = \frac{\lambda}{\nu}. \quad (2)$$

(1) et (2) montrent que le birapport "affine" (celui de [1], §. 16.1) de ces quatre points de l'espace affine A est

$$\xi = \frac{\overline{M_1M_3}}{\overline{M_1M_4}} : \frac{\overline{M_2M_3}}{\overline{M_2M_4}} = \frac{\mu}{\tau} : \frac{\lambda}{\nu}.$$

Calculons maintenant le birapport "projectif" $[M_1,M_2,M_3,M_4]$ en retournant à la Définition 3.11. Utilisons les coordonnées homogènes des points M_i dans le repère projectif (M_1,M_2,P) de la droite $\mathcal{D}$ (où $M_1=p\,(e_0)$, $M_2=p\,(e_1)$ et $P=p\,(e_0+e_1)$). Dans ce repère :

$$M_1 = (1:0) \ ; \quad M_2 = (0:1) \ ; \quad M_3 = (\lambda:\mu) \ ; \quad M_4 = (\nu:\tau) \ ;$$

donc

$$[M_1, M_2, M_3, M_4] = \frac{\begin{vmatrix} 1 & \lambda \\ 0 & \mu \end{vmatrix}}{\begin{vmatrix} 1 & \nu \\ 0 & \tau \end{vmatrix}} : \frac{\begin{vmatrix} 0 & \lambda \\ 1 & \mu \end{vmatrix}}{\begin{vmatrix} 0 & \nu \\ 1 & \tau \end{vmatrix}} = \frac{\mu}{\tau} : \frac{\lambda}{\nu} = \xi$$

et tout est pour le mieux dans le meilleur des mondes[6] ! Nous pouvons énoncer :

Théorème 3.28 *Le birapport de quatre point s'écrit à l'aide de mesures algébriques :*

$$[M_1, M_2, M_3, M_4] = \frac{\overline{M_1M_3}}{\overline{M_1M_4}} : \frac{\overline{M_2M_3}}{\overline{M_2M_4}}$$

dans n'importe quelle carte affine de l'espace projectif.

Que se passe-t-il lorsque le point M_4 de la FIG. 3.16 est envoyé à l'infini ? Autrement dit, que devient l'expression

$$\xi = \frac{\overline{M_1M_3}}{\overline{M_1M_4}} : \frac{\overline{M_2M_3}}{\overline{M_2M_4}}$$

lorsque M_4 est égal à $\infty_{\mathcal{D}}$ (où $\{\infty_{\mathcal{D}}\} = \mathcal{D} \cap H$) ?
Pour obtenir les coordonnées homogènes de $M_4 = \infty_{\mathcal{D}}$ dans le repère projectif de $\mathbb{P}\left(K^{n+1}\right)$ associé à $e = (e_0, e_1, ..., e_n)$, on remarque que

$$e_0 - e_1 \in \text{Vect}\,(e_0, e_1) \cap H,$$

donc que

$$M_4 = \infty_{\mathcal{D}} = p\,(e_0 - e_1) = (1 : -1 : 0 : ... : 0 : 0)$$

[6]La théorie très sérieuse du philosophe et mathématicien Gottfried Wilhelm Von Leibniz, pour laquelle "tout est pour le mieux dans le meilleur des mondes", a été mise à l'épreuve de la vie par Voltaire dans "*Candide ou l'optimisme*". La fin du roman est un morceau de bravoure dont on ne se lasse pas : "Toute la petite société entra dans ce louable dessein ; chacun se mit à exercer ses talents. La petite terre rapporta beaucoup. Cunégonde était, à la vérité, bien laide ; mais elle devint une excellente pâtissière ; Paquette broda ; la vieille eut soin du linge. Il n'y eut pas jusqu'à frère Giroflée qui ne rendît service ; il fut un très bon menuisier, et même devint honnête homme.

Et Pangloss disait quelquefois à Candide : tous les événements sont enchaînés dans le meilleur des mondes possibles ; car enfin si vous n'aviez pas été chassé d'un beau château à grands coups de pied dans le derrière pour l'amour de mademoiselle Cunégonde, si vous n'aviez pas été mis à l'inquisition, si vous n'aviez pas couru l'Amérique à pied, si vous n'aviez pas donné un bon coup d'épée au baron, si vous n'aviez pas perdu tous vos moutons du bon pays d'Eldorado, vous ne mangeriez pas ici des cédrats confits et des pistaches. Cela est bien dit, répondit Candide, mais il faut cultiver notre jardin."

(coordonnées homogènes dans $\mathbb{P}\left(K^{n+1}\right)$). Par conséquent :

$$[M_1, M_2, M_3, \infty_{\mathcal{D}}] = \frac{\begin{vmatrix} 1 & \lambda \\ 0 & \mu \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ 0 & -1 \end{vmatrix}} : \frac{\begin{vmatrix} 0 & \lambda \\ 1 & \mu \end{vmatrix}}{\begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix}} = -\frac{\mu}{\lambda} = \frac{\overline{M_1M_3}}{\overline{M_2M_3}}.$$

Ce rapport $\frac{\overline{M_1M_3}}{\overline{M_2M_3}}$ est parfois noté $[M_1, M_2, M_3]$, ce qui permet d'écrire

$$[M_1, M_2, M_3, \infty_{\mathcal{D}}] = [M_1, M_2, M_3] = \frac{\overline{M_1M_3}}{\overline{M_2M_3}}.$$

Avec cette notation, on peut aussi écrire

$$[M_1, M_2, M_3, M_4] = \frac{[M_1, M_2, M_3]}{[M_1, M_2, M_4]} = \frac{\overline{M_1M_3}}{\overline{M_1M_4}} : \frac{\overline{M_2M_3}}{\overline{M_2M_4}}$$

pour tous M_1, M_2, M_3, $M_4 \in \mathcal{D}_a$, dès que ces quatre points sont distincts.

3.10.6 Homographies laissant un hyperplan invariant

Nous avons besoin de quelques rappels de géométrie vectorielle :

Définition 3.12 *Soient F et G deux sous espaces vectoriels supplémentaires d'un espace vectoriel E, et $k \in K$. L'application*

$$\begin{array}{rccc} f: & E = F \oplus G & \to & E \\ & x = x_1 + x_2 & \mapsto & x_1 + kx_2 \end{array}$$

*est appelée **affinité vectorielle de base (ou par rapport à) F, de direction G et de rapport k.***

Définition 3.13 *Soient H un hyperplan de l'espace vectoriel E, l une forme linéaire telle que $H = \operatorname{Ker} l$, et h un vecteur de H. L'application*

$$\begin{array}{rccc} g: & E & \to & E \\ & x & \mapsto & x + l(x)\,h \end{array}$$

*est appelée **transvection vectorielle de base (ou par rapport à) H et de vecteur h.***

Le résultat important concernant les affinités et les transvections est le suivant (la version affine de ce Théorème est proposée en [1], Th. 57 mais ne nous intéresse pas ici) :

Théorème 3.29 *Soient E un espace vectoriel de dimension finie, et H un hyperplan de E. L'ensemble des automorphismes de E qui laissent H invariant point par point est formé des affinités de base H et de rapport $\neq 1$, et des transvections de base H et de vecteur non nul.*

Preuve : La condition est trivialement nécessaire. Montrons qu'elle est suffisante. Soit u un automorphisme de E tel que $\operatorname{Inv} u = \operatorname{Ker}(u - Id) = H$. On a :

$$\dim \operatorname{Im}(u - Id) = \dim E - \dim \operatorname{Ker}(u - Id) = \dim E - \dim H = 1,$$

donc $\operatorname{Im}(u - Id)$ est une droite que nous pouvons noter D.

De deux choses l'une :

• Si $D \not\subset H$, alors D est stable par u (en effet u et $u - Id$ commutent donc $u(u(x) - x) = (u - Id)(u(x)))$ donc $u|_D$ est une homothétie de rapport $k \neq 0$. Comme $E = H \oplus D$ et $u|_H = Id_H$, l'application u ne peut être que l'affinité de base H et de rapport k.

• Si $D \subset H$, on choisit un vecteur non nul D de D et l'on note $\varphi : D \to K$ l'isomorphisme qui à λD associe λ. On a le diagramme

$$\begin{array}{ccccc} E & \overset{u-Id}{\longrightarrow} & D & \overset{\varphi}{\longrightarrow} & K \\ & & \lambda \overrightarrow{d} & \longmapsto & \lambda. \end{array}$$

L'application $l = \varphi \circ (u - Id)$ est une forme linéaire de noyau H, et pour tout $x \in E$,

$$(u - Id)(x) = \varphi^{-1}(l(x)) \quad \Rightarrow \quad u(x) = x + l(x)\,\overrightarrow{d}.$$

Cela montre que u est une transvection de base H. ■

On peut maintenant énoncer un résultat remarquable :

Théorème 3.30 *Soient $\mathbb{P}(E)$ un espace projectif de dimension n et H un hyperplan de $\mathbb{P}(E)$. Soit H_1 l'espace vectoriel associé à H (i.e. $H = p(H_1)$). Soit T l'ensemble des homographies de $\mathbb{P}(E)$ qui admettent H comme ensemble de points invariants auquel on adjoint l'identité.*

1) T est formé des homographies $P(u)$ associées aux transvections u de base H_1.

2) T est un groupe isomorphe au groupe additif de l'espace vectoriel H_1,

3) T et H_1 agissent simplement et transitivement sur $\mathbb{P}(E) \setminus H$, de sorte que le complémentaire d'un hyperplan projectif soit toujours structuré en espace affine.

Preuve : 1) Soit $P(u)$ une homographie admettant $H = p(H_1)$ comme ensemble de points invariants. On a

$$\forall x \in H_1 \quad u(x) \text{ colinéaire à } x,$$

donc $u|_{H_1}$ est une homothétie, et il existe $k \in K^*$ tel que

$$\forall x \in H_1 \quad u(x) = kx.$$

Quitte à modifier u, on peut supposer $k = 1$. Alors $\operatorname{Inv} u = H_1$ et le Théorème 3.29 montre que u est soit une affinité, soit une transvection. Cela ne peut être une affinité, sinon il existerait un vecteur $a \notin H_1$ et un scalaire $k \in K$ tels que $u(a) = ka$, et l'on aurait $P(u)(\dot{a}) = \dot{a}$ avec $\dot{a} \notin H$, ce qui est absurde. C'est donc une transvection : il existe $h_u \in H_1$ tel que

$$\forall x \in E \quad u(x) = x + l(x)\, h_u.$$

2) Il est facile de vérifier que $(T, \circ)$ est un groupe. L'application

$$\begin{array}{rccc} \Psi : & T & \to & H_1 \\ & P(u) & \mapsto & h_u \end{array}$$

est bijective, et c'est un isomorphisme de $(T, \circ)$ sur $(H_1, +)$ puisque, si

$$\left\{ \begin{array}{l} u(x) = x + l(x)\, h_u \\ v(x) = x + l(x)\, h_v, \end{array} \right.$$

alors, pour tout $x \in E$,

$$\begin{array}{rclclcl} (v \circ u)(x) & = & v(u(x)) & = & v(x + l(x)\, h_u) & = & v(x) + l(x)\, v(h_u) \\ & & & & & = & x + l(x)\, h_v + l(x)\, h_u \end{array}$$

soit $(v \circ u)(x) = x + l(x)(h_v + h_u)$.

3) On vérifie que le groupe T opère simplement et transitivement sur $P \backslash H$ par l'application

$$\begin{array}{rccc} \Phi : & (\mathbb{P}(E) \backslash H) \times T & \to & \mathbb{P}(E) \backslash H \\ & (\dot{a}, P(u)) & \mapsto & P(u)(\dot{a}) = \dot{\overline{a + l(a)\, h_u}}. \end{array} \quad \blacksquare$$

Remarque : Le Théorème 3.6 nous a déjà montré que le complémentaire d'un hyperplan projectif est structuré en espace affine, mais la nouvelle preuve proposée au Théorème 3.30 est purement géométrique, c'est-à-dire sans appel à des coordonnées projectives.

Bibliographie

[1] D.-J. Mercier, Cours de géométrie, préparation au CAPES et à l'agrégation, Publibook, 2004.

[2] D.-J. Mercier, L'épreuve d'exposé au CAPES mathématiques, 14 leçons rédigées et commentées,Vol. I, Publibook, 2004.

[3] J.-C. Sidler, Géométrie projective, Cours, exercices et problèmes corrigés, Dunod, 2ème édition, 1993.

[4] X. Siefridt, Espaces projectifs, RMS 1989-90, n^o**8,** 1990.

[5] P. Samuel, Géométrie projective, PUF, 1986.

Chapitre 4

Sur les polyèdres eulériens

Polyèdres eulériens et solides pathologiques.

(Dany-Jack Mercier[1])

Résumé : La formule $S - A + F = 2$ concernant des polyèdres, proposée par Leonhard Euler en 1750 sans qu'il soit capable d'en donner une preuve rigoureuse, attendra 1794 pour être démontrée par Adrien Marie Legendre. Dans cet article, nous voulons préciser son champ d'application, imaginer des contre-exemples, et retourner sur deux preuves classiques bien jolies : celle de Cauchy et celle utilisant la même relation $s - a + f = 2$ vraie pour des graphes connexes. Cette petite incursion dans le monde des polyèdres s'achèvera avec la preuve de l'existence de seulement cinq polyèdres réguliers, encore appelés "solides platoniciens" ou "polyèdres parfaits".

4.1 Polyèdres eulériens

C'est en 1750 que Leonhard Euler propose une formule générale concernant les polyèdres. L'idée d'Euler est de comparer les polyèdres à leurs analogues dans le plan : les polygones. Cette analogie ne peut malheureusement pas être menée bien loin car trop de différences séparent ces deux objets d'étude. Ainsi, si un polygone possède autant de côtés que d'angles, il n'en est plus de même d'un polyèdre dont le nombre de faces ne correspond plus au nombre d'angles solides qu'il détermine.

[1] IUFM de Guadeloupe, dany-jack.mercier@univ-ag.fr.

Euler introduit le terme "d'arêtes" (latus en latin, qui a donné lattice en anglais, et qui signifie alors "grillage" ou "treillis") et vérifie la règle

$$S - A + F = 2$$

sur de nombreux exemples.

Dans cette formule, S désigne le nombre de sommets d'un polyèdre, A le nombre de ses arêtes et F le nombre de ses faces. Euler est incapable de proposer une démonstration rigoureuse de cette formule, et la preuve qu'il annonce un an plus tard contient un défaut embarrassant.

Il faudra attendre le traité d'Adrien Marie Legendre, intitulé "Eléments de géométrie" et paru en 1794 en pleine révolution française, pour lire une démonstration satisfaisante de la formule d'Euler. Cette démonstration, qui fait appel à des notions de trigonométrie sphérique, reste assez éloignée de l'énoncé initial...
La preuve proposée par Augustin Louis Cauchy en 1811 est simple et élégante, et c'est évidemment celle que l'on décrit dans la Section suivante. Cette preuve possède en outre le mérite de s'appliquer à des polyèdres beaucoup plus généraux que ceux envisagés par Legendre, et utilise les parties constitutives d'un polygone (faces, arêtes, sommets) de manière pertinente.

4.2 Preuve de Cauchy (1811)

La preuve se fait en plusieurs étapes visualisées, dans la cas d'un cube, sur la FIG. 4.1.

1. De (a) vers (b) : On choisit une face du polyèdre comme base, puis on transporte tous les éléments sur cette base en conservant leur nombre et leur position relative. On raisonne maintenant sur une figure plane (un graphe du plan) qui possède une face de moins que le polyèdre. Pour un tel graphe, il s'agit de montrer la formule $S - A + F = 1$.

2. De (b) vers (c) : On rajoute au réseau autant de diagonales que nécessaire pour que toutes ses faces soient triangulaires, autrement dit on triangule la figure.

3. De (c) jusqu'à (f) : On retire un à un les triangles du réseau en partant de l'extérieur. De cette façon, on supprime soit un sommet, une face et deux arêtes, soit une arête et une face. Dans les deux cas, la quantité $S - A + F$ demeure inchangée.

4. Conclusion : La figure (f) de comporte plus qu'un seul triangle pour lequel la formule $S - A + F = 1$ est évidente.

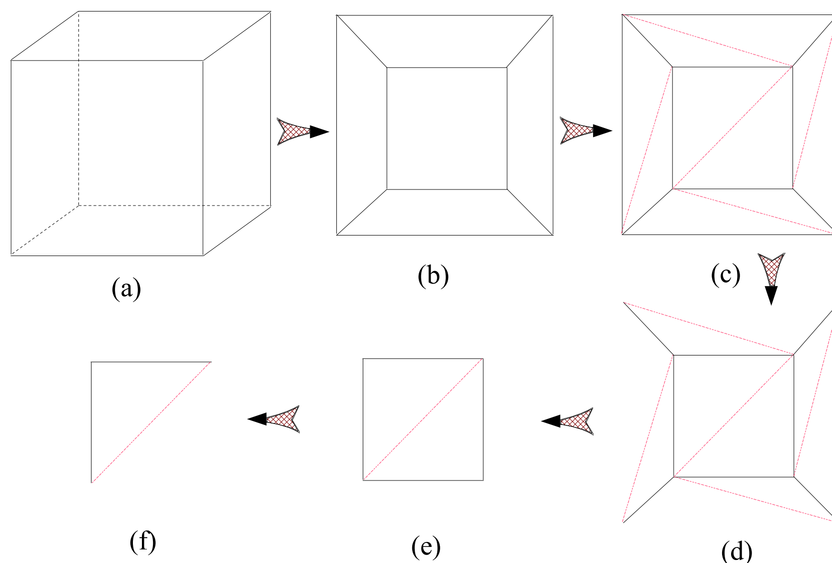


FIG. 4.1 – Preuve de Cauchy (le cas du cube)

4.3 Les solides pathologiques de Lhuilier

Deux ans après la preuve de Cauchy, le mathématicien genevois Simon Lhuilier met en garde ses collègues contre toute généralisation abusive de la formule $S - A + F = 2$, et exhibe un certain nombre de solides "pathologiques".
La difficulté réside sans doute déjà dans la définition que l'on peut donner du mot "polyèdre". Ainsi, un polyèdre est-il :

- un solide plein limité par des faces planes ?
- ou plutôt une surface constituée d'un système de polygones ?

Des définitions plus ou moins originales sont données, parfois seulement pour éliminer certains solides pathologiques de Lhuilier [4]. Les deux définitions suivantes nous seront utiles :

Définition 4.1 *Un polyèdre est* ***eulérien*** *s'il vérifie $S - A + F = 2$.*

Définition 4.2 *Un* ***polyèdre convexe (fermé)*** *est une partie non vide et bornée de l'espace obtenue comme l'intersection d'un nombre fini de demi-espaces fermés.*
Un point est un ***sommet*** *du polyèdre s'il appartient à celui-ci et à l'intersection d'au moins trois frontières de demi-espaces fermés définissant le polyèdre.*

Une ***arête*** *du polyèdre est un segment joignant deux sommets et inclus dans l'une des frontières des demi-espaces fermés définissant le polyèdre.*

On imagine bien que tout polyèdre convexe au sens de la définition précédente est eulérien. Mais dans un cadre plus général, quels sont les contre-exemples de Lhuilier ?
On dénombre trois grandes catégories d'exceptions.

▶ **Première catégorie : les polyèdres à structures multiples**.

Un petit cube posé sur une face d'un grand cube (FIG. 4.2.a) possède 16 sommets, 24 arêtes et 11 faces, donc vérifie $S - A + F = 3$.
On peut aussi imaginer que le petit cube creuse l'une des faces du grand cube en formant une dépression, ce qui ne change rien au calcul précédent.

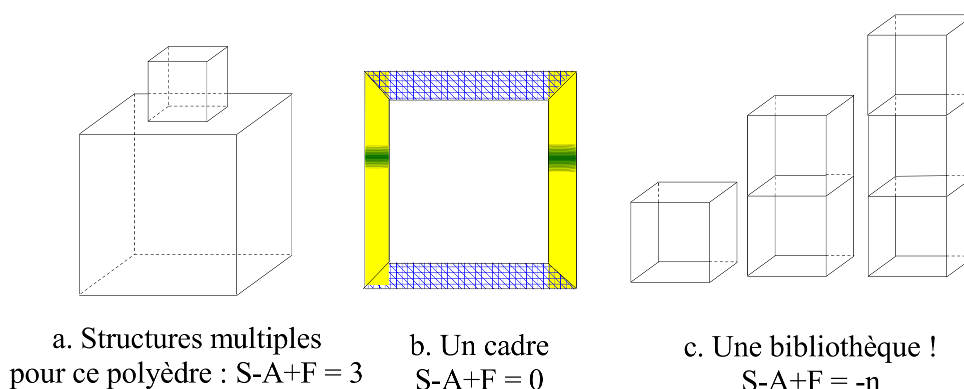


FIG. 4.2 – Structures multiples et tunnels

▶ **Deuxième catégorie : les polyèdres à tunnels**.

Un simple cadre (penser à un cadre en bois entourant une peinture) vérifie $S - A + F = 0$ (FIG. 4.2.b). On peut aussi simplement empiler n cubes-tunnels (c'est-à-dire des cubes privés de deux faces opposées) pour obtenir des polyèdres non eulériens pour lesquels $S - A + F = -n$ (FIG. 4.2.c).

▶ **Troisième catégorie : les polyèdres à cavités**.

Un petit cube imbriqué dans un grand nous donne un contre-exemple simple (FIG. 4.3.a). Bien entendu, il ne s'agit pas d'un polyèdre convexe, mais c'est à ce prix qu'il n'est plus eulérien et satisfait $S - A + F = 4$.
C'est en observant une collection minéralogique montrant des cristaux, et en particulier un cristal opaque pris dans un cristal translucide, que Lhuilier eut l'idée de ce contre-exemple.

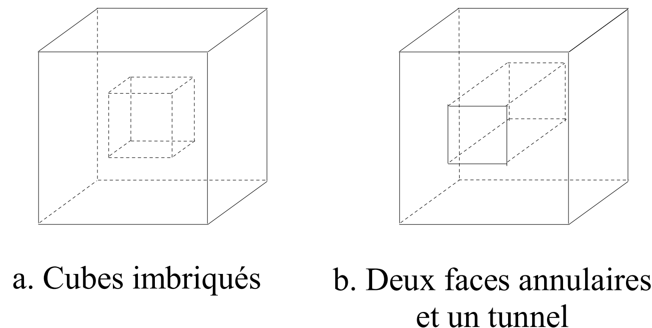


FIG. 4.3 – Toujours avec des cubes...

4.4 Une caractérisation des polyèdres eulériens

La preuve de Cauchy ne peut s'appliquer qu'aux polyèdres qui vérifient les trois propriétés suivantes :

> **C1**. Une des faces étant choisie comme base, toutes les autres doivent pouvoir être transportées sur elle sans déchirure ni duplicature,
>
> **C2**. Le réseau obtenu doit pouvoir être triangulé,
>
> **C3**. Les triangles obtenus doivent pouvoir être retirés comme prévu.

De façon évidente, tout polyèdre satisfaisant les propriétés **C1** à **C3** est eulérien, et la question est de savoir si la réciproque est vraie. Il n'en est rien, comme le montre le polyèdre de la FIG. 4.3.b qui est eulérien sans vérifier **C1** à **C3**.

C'est Christian Von Staudt qui énonce le premier une caractérisation des polyèdres eulériens. Dans sa "Géométrie de position", publiée en 1847, il démontre qu'un polyèdre est eulérien si et seulement si :

a) Chacun de ses sommets peut être joint à tout autre par une ligne formée d'arêtes,

b) Toute ligne (fermée) de ce genre passant au plus une fois par un même sommet sépare la surface en 2 parties.

La FIG. 4.4 montre quelques lignes fermées du cube satisfaisant la propriété b).

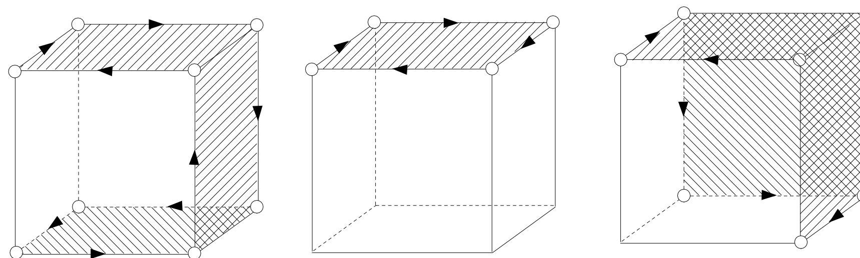


FIG. 4.4 – Caractérisation d'un polyèdre eulérien

4.5 Relation d'Euler pour des graphes connexes

A. Le langage des graphes

Le langage des graphes est adapté à l'étude des polyèdres, la démonstration de Cauchy ayant la particularité de transformer un problème spatial en un problème dans le plan ne faisant intervenir que des sommets, des arêtes et des faces.
Rappelons qu'un **graphe** (encore appelé **réseau**) est un schéma formé de points et d'**arêtes** qui relient certains de ces points. Les points, supposés en nombre fini, s'appellent des **sommets** ou des **noeuds**. L'**ordre** du graphe est le nombre de ses sommets, le **degré** (ou la **puissance**) d'un sommet est le nombre d'arêtes qui admettent ce sommet pour extrémité.

On dit que deux sommets liés par au moins une arête sont **adjacents** (ou **voisins**), qu'une **chaîne** (ou **chemin**) est une liste ordonnée d'arêtes telle que l'extrémité de chacune (sauf la dernière) soit l'origine de la suivante, et qu'un chemin est **fermé** si son origine et son extrémité sont confondues. Enfin :

- Un **cycle** est un chemin fermé composé d'arêtes toutes distinctes,
- Un graphe est **connexe** si deux sommets distincts quelconques sont toujours reliés par au moins une chaîne,
- Un **point terminal** est un sommet où n'aboutit qu'une seule arête.

(Que faut-il savoir sur les graphes pour le CAPES ? voir [2], Chap. 2)

Avec ces définitions :

Lemme 4.1 *Un graphe connexe qui ne contient aucun cycle possède au moins un point terminal.*

Preuve : Un graphe connexe sans aucun point terminal permet la construction suivante. Partant d'un sommet A_0 quelconque, on choisit une arête d'extrémité

A_0 et A_1. Le sommet A_1 n'est pas terminal, donc il existe un sommet A_2 autre que A_0 et lié à A_0 par une arête. On obtient une chaîne $A_0A_1A_2$, puis l'on continue de proche en proche. Le nombre de sommets étant fini, il existera un moment où l'on obtiendra un sommet déjà présent dans la chaîne : on s'arrête alors pour exhiber un cycle $A_0A_1...A_{l-1}$. ■

Armé de ce Lemme, on peut démontrer une "relation d'Euler" concernant des graphes. Le lien avec la relation d'Euler sur les polyèdres apparaîtra juste après.

Théorème 4.1 *Soit G un graphe connexe (non vide) dessiné sur une sphère S de l'espace affine euclidien de dimension 3. Soit f le nombre de composantes connexes de $S\backslash G$ dans S. Le nombre a d'arêtes et le nombre s de sommets de G vérifient la relation d'Euler $s-a+f=2$.*

Preuve : On raisonne par récurrence sur le nombre d'arêtes de G. Si $a=0$, $s=1$, $f=1$ et la formule est triviale. Supposons la propriété démontrée jusqu'au rang $a-1$, et considérons un graphe connexe G possédant a arêtes. De deux choses l'une :

✓ **Premier cas** : Il existe un cycle $A_0...A_{l-1}A_0$ dans G. Dans ce cas, on supprime une arête de ce cycle, par exemple $[A_0A_1]$, pour obtenir un graphe G' à $a-1$ arêtes, s sommets et $f-1$ composantes connexes. L'hypothèse récurrente s'applique à G', et donne

$$s-(a-1)+(f-1)=2,$$

soit $s-a+f=2$.

✓ **Second cas** : G ne contient aucun cycle. Le Lemme précédent montre l'existence d'un sommet terminal A de G. Le graphe G' obtenu en supprimant ce sommet et l'unique arête qui y aboutit possède $a-1$ arêtes, $s-1$ sommets et toujours f composantes connexes. L'hypothèse récurrente donne

$$(s-1)-(a-1)+f=2,$$

soit encore $s-a+f=2$. ■

B. Une autre preuve de la relation d'Euler pour les polyèdres

Considérons maintenant un polyèdre convexe (Définition 4.2). De façon intuitive, un tel polyèdre ne "possède aucun trou", et on peut le "gonfler" jusqu'à obtenir une sphère. On montre alors :

Corollaire 4.1 *Tout polyèdre convexe (non aplati) est eulérien, autrement dit possède S sommets, A arêtes, et F faces satisfaisant la relation $S-A+F=2$.*

Preuve : Un polyèdre convexe non aplati $\mathcal{P}$ contient nécessairement quatre sommets A, B, C, D non coplanaires. L'isobarycentre G de ces quatre sommets appartient à l'intérieur du tétraèdre $T = ABCD$. Puisque $\mathcal{P}$ est convexe, il contient T, et G est un point de l'intérieur topologique de $\mathcal{P}$. On peut donc trouver un réel strictement positif r tel que la sphère S de centre G et de rayon r soit incluse dans l'intérieur de $\mathcal{P}$. L'application

$$\begin{array}{cccc} \xi : & \partial\mathcal{P} & \to & S \\ & M & \mapsto & \xi(M) \end{array}$$

qui au point M de la frontière $\partial\mathcal{P}$ de $\mathcal{P}$ associe l'unique point $N = \xi(M)$ intersection de $[GM]$ et S, est bijective. Elle est aussi continue puisque

$$\overrightarrow{GN} = \frac{r}{||\overrightarrow{GM}||}\overrightarrow{GM},$$

c'est donc un homéomorphisme de $\partial\mathcal{P}$ sur S. Les images par ξ des sommets et des arêtes de $\mathcal{P}$ forment un graphe connexe dessiné sur la sphère S, et l'on peut appliquer le Théorème 4.1. ■

Cette preuve de la relation d'Euler nous montre l'importance des propriétés topologiques des ensembles sur lesquels on dessine. Des graphes dessinés dans un espace affine, sur une sphère, ou encore sur un tore, n'ont pas la même "valeur". On peut par exemple penser à la relation d'homotopie des chemins continus fermés dessinés dans un espace topologique : deux chemins continus fermés (des "lacets") sont homotopes s'il existe une déformation continue du premier qui amène sur le second. Si un lacet L dessiné sur une sphère est toujours homotope à un point, il n'en est pas de même d'un lacet dessiné sur un tore (il suffit pour cela que ce lacet "entoure" le "trou" du tore). Bref, ces notions nous mènent directement sur celle de classe d'homotopie dans un espace topologique.
La preuve précédente fonctionne avec des polyèdres convexes, mais pourrait être adaptée à des polyèdres "pathologiques" en décidant de projeter arêtes et sommets de celui-ci sur un espace topologique adapté (par exemple un tore si notre polyèdre est formé de deux cubes imbriqués, comme à la FIG. 4.3.a). Les notions de connexité par arcs et d'homotopie jouent un rôle crucial dans l'étude de « polyèdres généraux ».

C. Une conséquence importante : les 5 solides platoniciens

La relation d'Euler permet de montrer facilement qu'il n'existe que 5 polyèdres réguliers convexes. Ces polyèdres « parfaits » apparaissent dans les discours de Platon, et méritent à ce titre d'être qualifiés de « platoniciens »[2].

[2] Socrate : Dis-moi, Théétète, ce qu'est un polyèdre.

Définition 4.3 *Un polyèdre convexe non aplati est **régulier** si :*
a) Sur chacun des sommets aboutissent le même nombre p d'arêtes ($p \geq 3$),
b) Chacune des faces possède le même nombre q de côtés ($q \geq 3$).

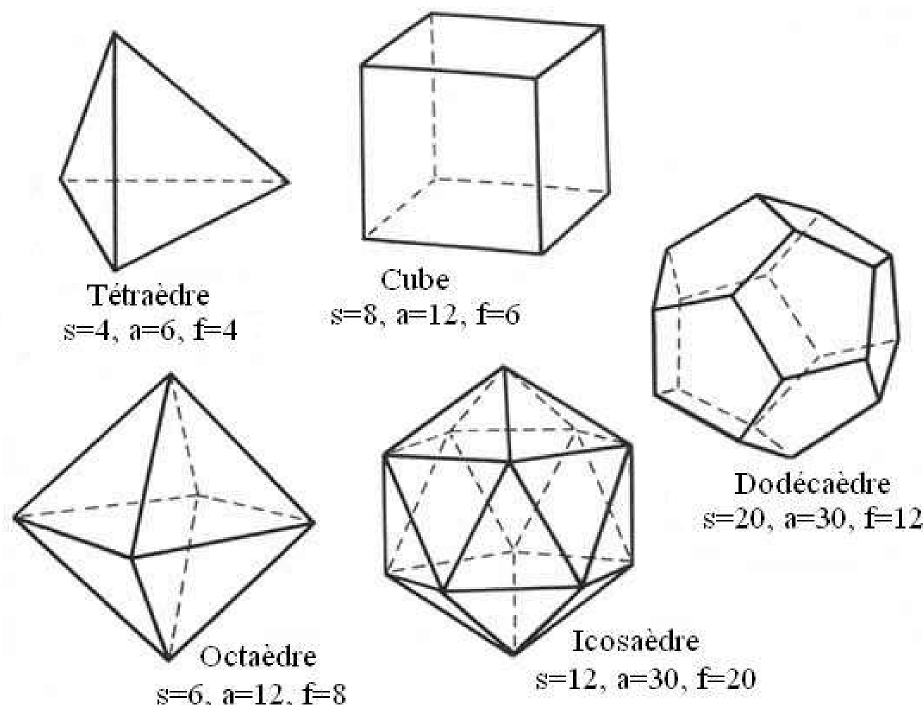


FIG. 4.5 – Les 5 solides platoniciens

Théorème 4.2 *Il n'existe que* 5 *polyèdres régulier (ce sont les* 5 *solides platoniciens représentés à la FIG. 4.5).*

Preuve : Soit $\mathcal{P}$ un polyèdre régulier. Une arête est le côté de deux faces, donc $Fq = 2A$. Puisque p arêtes aboutissent sur chacun des sommets, on devrait

Théétète : Eh bien, le tétraèdre, le cube, l'octaèdre, le dodécaèdre et l'icosaèdre sont des polyèdres...

Socrate : Ce sont là des exemples, je te demande ce qu'est le polyèdre...

La question posée par Socrate dans ce dialogue reste primordiale ! Définir un polyèdre n'est pas facile, et chercher un échappatoire dans l'énumération des polyèdres que l'on connaît ne suffit pas à définir cet objet d'étude. Les premiers dialogues de Platon montrent souvent un élève hésitant, mais plein de bonne volonté, qui échoue à définir l'essence du Beau, de la Vertu, de la Justice, ... et ne fait que proposer des exemples.

dénombrer Sp arêtes, mais en procédant ainsi on compte chacune des arêtes deux fois (une fois pour chacun de ses deux sommets), donc $Sp = 2A$. On est ainsi amené à résoudre le système

$$\begin{cases} S - A + F = 2 \\ Fq = 2A \\ Sp = 2A. \end{cases}$$

On obtient

$$S = \frac{4q}{2p + 2q - pq}, \quad A = \frac{2pq}{2p + 2q - pq}, \quad F = \frac{4p}{2p + 2q - pq}.$$

On est alors amené à chercher tous les couples d'entiers (p, q) qui vérifient les trois inégalités $2p + 2q - pq > 0$, $p \geq 3$ et $q \geq 3$. La première condition, qui s'écrit

$$(p - 2)(q - 2) < 4,$$

impose à p et q d'être petits, et nous donne seulement 5 couples d'entiers possibles : $(3, 3)$, $(3, 4)$, $(3, 5)$, $(4, 3)$ et $(5, 3)$. On vérifie alors que chacun de ces couples permet le calcul d'un triplet (S, A, F) correspondant à l'un des cinq polyèdres platoniciens de la FIG. 4.5. ■

Bibliographie

[1] D.-J. Mercier, Cours de géométrie, préparation au CAPES et à l'agrégation, Publibook, 2004.

[2] D.-J. Mercier, L'épreuve d'exposé au CAPES mathématiques, 14 leçons rédigées et commentées, Vol. I, Publibook, 2007.

[3] T. Eveilleau, page web "Les solides de Platon" en http ://pagesperso-orange.fr/therese.eveilleau%20/pages/truc_mat/textes/platon.htm.

[4] G. Orvas, Des solides pathologiques, Les Cahiers de Science & Vie n°59, octobre 2000, pp. 60-63.

Chapitre 5

Outils élémentaires de l'analyse

Outils élémentaires de l'analyse.
(Robert Rolland[1])

Résumé : L'analyse est un secteur très large de l'activité mathématique. Développée à partir du XVIIe siècle, elle a été essentiellement centrée sur le calcul infinitésimal jusqu'à la fin du XIXe siècle. Cette partie s'occupe des nombres et des fonctions en tant que tels, des bonnes approximations de ces objets et des outils de dérivation et d'intégration qui pemettent d'opérer sur eux et même éventuellement de les définir.

À partir du XXe siècle, l'analyse fonctionnelle, qui considère qu'une fonction est un point d'un ensemble de fonctions, avec tous les aspects géométriques qui s'y rattachent, s'est developpée, principalement pour la résolution des problèmes aux limites des équations différentielles et intégrales.

Nous présentons ici un texte libre qui donne quelques idées élémentaires sur les outils de base et leurs utilisations dans le cadre du calcul infinitésimal. Nous insistons sur l'emploi simultané de techniques différentielles et de techniques intégrales, le tout relié bien entendu par le théorème fondamental du calcul différentiel et intégral.

L'aspect analyse fonctionnelle n'est pas abordé, tout au moins comme objet central d'étude.

[1] Institut de Mathématiques de Luminy, robert.rolland@acrypta.fr.

5.1 Introduction

Dans ces notes, nous nous préoccupons d'une partie particulière de l'analyse, le **calcul infinitésimal**. Celui-ci, qui s'appuie sur les outils classiques que sont la dérivation et l'intégration, c'est-à-dire sur le **calcul différentiel** et le **calcul intégral**, se préoccupe essentiellement d'approcher des nombres et des fonctions. En reprenant une formule employée par Jean Dieudonné dans son livre *Calcul infinitésimal*, le calcul infinitésimal peut se résumer en trois mots : **majorer, minorer, approcher**.

Au début du XXe siècle, lors du Congrés International de Mathématiques de 1897, Jacques Hadamard énonce clairement **qu'il faut considérer les fonctions comme des points dans des ensembles de fonctions**. Ceci est un constat de l'état de la recherche en analyse à cette époque et des directions qu'elle prend alors. Désormais, va se développer, avec en vue l'étude des équations intégrales et des problèmes aux limites pour les équations différentielles, la théorie des espaces fonctionnels, mettant l'accent sur une géométrie des espaces de dimension infinie. On assiste là, au développement de **l'analyse fonctionnelle**. Nous ne parlerons pas de cet aspect, tout au moins en tant que théorie, nous contentant ici d'exprimer les résultats anciens du calcul infinitésimal, quand ceci s'impose, dans le cadre de l'analyse actuelle.

En définitive, nous parlerons de nombres, fonctions, suites, séries, dérivées et intégrales en relation avec des problèmes d'approximation et d'interpolation. Nous évaluerons des différences, à l'aide de majorations, minorations, des ordres de grandeur, nous étudierons des comportements asymptotiques.

Le texte que nous proposons, **n'est pas un cours** suivi de calcul infinitésimal, mais regroupe plutôt quelques notes un peu disparates, écrites au fil de la plume, qui soulignent quelques idées de base, agrémentées d'exemples, qui me semblent importantes pour celui qui doit aborder des exercices et problèmes, ou préparer des cours sur ce sujet.

De ce fait, ce document n'a bien entendu pas vocation à être exhaustif, ni à aborder des résultats fins d'analyse. Pour tout cela, le lecteur pourra se référer aux nombreux cours de « Calculus » disponibles.

5.2 Dichotomie

Je ne voudrais pas commencer une présentation des outils et des méthodes élémentaires de l'analyse, sans citer le plus simple d'entre eux : la dichotomie. C'est un moyen performant pour encadrer des nombres, notamment des solu-

tions d'équations du type $f(x) = 0$, qui s'apparente aux stratégies employées en informatique sous le nom de « diviser pour régner ».

Nous partons d'une équation :

$$f(x) = 0,$$

où f est une fonction continue, dont nous supposons que nous avons isolé le zéro c que nous voulons calculer. C'est-à-dire que nous avons trouvé un intervalle [a,b] tel que :

$$\begin{cases} c \in]a,b[, \\ \forall x \in [a,b] \setminus \{c\}, f(x) \neq 0. \end{cases}$$

Nous supposerons en outre que f change de signe en c , donc que :

$$f(a)f(b) < 0.$$

La méthode de calcul de c par dichotomie est très simple : on coupe l'intervalle sur lequel on travaille en deux et on teste sur lequel des deux sous-intervalles obtenus se trouve c. On réitère ensuite le procédé à partir du sous-intervalle déterminé. Après n itérations, on localise le nombre c sur un intervalle de longueur $(b-a)/2^n$. Voici les détails :

On fixe $\epsilon > 0$,

```
A := a ;
B := b ;
E := ε ;
tant que  B − A > E faire
    C := (A + B)/2 ;
    si f(C) = 0 alors
        retourner C ;
        sortir ;
    finsi ;
    si f(A)f(C) < 0 alors
        B := C ;
    sinon
        A := C ;
    finsi ;
fintq ;
retourner C ;
```

Cette méthode ne s'applique pas lorsque f s'annule sans changer de signe, ou alors s'annule en des points si proches, qu'il n'est pas possible, numériquement, d'isoler un zéro.

5.3 Inégalité des accroissements finis

Soit f une fonction définie sur un intervalle (a, b). Devoir évaluer une quantité du type $|f(x) - f(y)|$ est un problème très fréquent, notamment lors des calculs d'erreurs en physique. Un cas très favorable est celui d'une fonction K-lipschitzienne.

Définition 5.1 *Soit $K > 0$ une constante. Nous dirons que la fonction f définie sur (a, b) est K-lipschitzienne, si pour tout x et tout y de l'intervalle (a, b) on a la majoration :*

$$|f(x) - f(y)| \leq K|x - y|.$$

Nous dirons que f est lipschitzienne, s'il existe une constante $K > 0$ telle que f soit K-lipschitzienne. On dira aussi : f est lipschitzienne de rapport K.

Quand on arrive à montrer qu'une fonction est lipschitzienne, il en découle de bonnes propriétés. En effet la fonction est alors uniformément continue sur (a, b).
Obtenir de telles majorations, tout au moins sur une partie de $\mathbb{R}$, peut se faire facilement par calcul algébrique lorsque la fonction s'exprime elle même simplement sous forme algébrique. Par exemple si f est une fonction polynomiale, ou une fraction rationnelle, ou une racine carrée.

Exemple 1 : la fonction $f(x) = x^2$ est lipschitzienne sur $[0, 1]$. En effet :

$$f(x) - f(y) = x^2 - y^2 = (x - y)(x + y).$$

Donc :

$$|f(x) - f(y)| \leq 2|x - y|.$$

Cette fonction n'est pas lipschitzienne sur $[0, +\infty[$. D'ailleurs une fonction uniformément continue sur $[0, +\infty[$ est majorée par une fonction affine, ce qui n'est pas le cas de x^2.

Exemple 2 : la fonction $f(x) = \frac{x+1}{x-1}$ est lipschitzienne sur $[-1, 1/2]$. En effet :

$$f(x) - f(y) = \frac{x+1}{x-1} - \frac{y+1}{y-1} = 2\frac{y-x}{(x-1)(y-1)}.$$

Donc :

$$|f(x) - f(y)| \leq 2\frac{|y-x|}{|(x-1)(y-1)|},$$

d'où :

$$|f(x) - f(y)| \leq 8|y - x|.$$

Exemple 3 : la fonction $f(x) = \sqrt{x}$ est lipschitzienne sur $[1/2, 1]$. En effet :

$$f(x) - f(y) = \sqrt{x} - \sqrt{y} = \frac{x - y}{\sqrt{x} + \sqrt{y}}.$$

Donc :

$$|f(x) - f(y)| \leq \frac{\sqrt{2}}{2}|x - y|.$$

Il n'en est pas de même quand f est une fonction plus compliquée. Dans ce dernier cas il faut mettre en place un outil plus complexe : le théorème (ou inégalité) des accroissements finis.
Une version immédiate de ce théorème se démontre avec le théorème fondamental du calcul différentiel et intégral, dans le cas d'une bonne fonction f dérivable sur un segment $[a, b]$ et dont la dérivée est intégrable (pour nous ce sera ici au sens de Riemann, donc en particulier on supposera $f'(t)$ bornée sur $[a, b]$). On a alors pour tout x et tout y de $[a, b]$:

$$f(y) - f(x) = \int_x^y f'(t)dt. \tag{5.1}$$

En posant :

$$M = sup_{t\in[a,b]}|f'(t)|dt,$$

on obtient la majoration uniforme :

$$|f(x) - f(y)| \leq M|x - y|.$$

La forme plus générale suivante, a une **expression géométrique** très simple et très intuitive.

Théorème 5.1 (Théorème des accroissements finis) *Soit f une fonction continue sur le segment $[a, b]$, dérivable sur l'intervalle ouvert $]a, b[$. Alors il existe un point $c \in]a, b[$ tel que :*

$$f(b) - f(a) = f'(c)(b - a).$$

Autrement dit il existe un point c intérieur au segment où la tangente est parallèle à la corde joignant l'origine à l'extrémité de l'arc considéré (cf. figure 5.1).

Preuve. Il suffit de se ramener au cas où $f(a) = f(b)$, cas dans lequel on doit montrer l'existence d'un point c intérieur tel que $f'(c) = 0$. Dans ce cas le

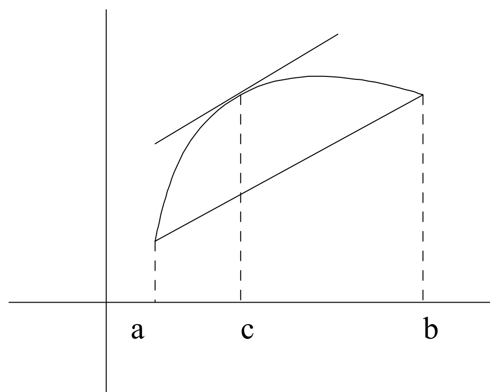


FIG. 5.1 – Théorème des acroissements finis

théorème s'appelle théorème de Rolle. Puis on applique le théorème de Rolle à la fonction :

$$g(x) = f(x) - \frac{f(b) - f(a)}{b - a}x,$$

qui en vérifie bien les hypothèses, ce qui nous donne directement le résultat. Supposons donc en outre, $f(a) = f(b)$ et montrons l'existence d'un point c tel que $f'(c) = 0$. Si la fonction est nulle sur $[a, b]$ le résultat a lieu. Sinon, il existe un point intérieur c où la fonction atteint un extremum (f est continue sur un compact). Supposons par exemple qu'au point c la fonction soit maximale. Alors pour tout $a \leq x < c$, on a

$$\frac{f(x) - f(c)}{x - c} \geq 0,$$

ce qui implique, par passage à la limite en c, que $f'(c) \geq 0$ (par hypothèse la dérivée existe en tout point intérieur). Par ailleurs, pour tout $c < x \leq b$ on a

$$\frac{f(x) - f(c)}{x - c} \leq 0,$$

ce qui implique cette fois que $f'(c) \leq 0$. On en déduit que $f'(c) = 0$. ■

Cette forme géométrique conduit directement à l'inégalité des accroissements finis, dans le cas où on peut encadrer la dérivée f' sur l'ouvert $]a, b[$.

Théorème 5.2 (Inégalité des accroissements finis) *Soit f une fonction continue sur un segment $[a, b]$ dérivable sur l'ouvert $]a, b[$ et telle qu'il existe deux constantes $M \geq m$ telles que pour tout $x \in]a, b[$ on ait :*

$$m \leq f'(x) \leq M.$$

Alors,

$$m(b-a) \leq f(b) - f(a) \leq M(b-a).$$

En posant $K = max(|m|, |M|)$*, on obtient :*

$$|f(b) - f(a)| \leq K|b-a|.$$

En conclusion, remarquons dès à présent, que lorsqu'on peut employer l'outil très puissant qu'est l'intégration (voir équation (5.1)), on obtient très rapidement et à moindre frais une majoration de $|f(y) - f(x)|$. Cependant, dans des cas plus particuliers (et certainement plus rares), on doit rester dans le cadre du calcul différentiel sans passer par le calcul intégral. C'est le cas ici, pour la preuve du théorème des accroissements finis (voir le théorème 5.1) supposant uniquement l'existence de la dérivée sur l'ouvert $]a, b[$. Si on sort du cadre de l'interprétation géométrique (qui a tout de même ici un intérêt certain) pour se plonger dans le calcul infinitésimal, il arrive un moment où il faut bien majorer. Et là, la marge de généralité, que semble avoir la fonction f, se réduit notablement. Nous retrouverons cette même question quand nous verrons la formule de Taylor.

5.4 Point fixe, méthode de Newton

5.4.1 Approximations successives, point fixe

Avant de présenter la méthode de Newton, disons quelques mots sur la méthode générale des approximations successives. Soit f une fonction de classe C^1 définie sur un intervalle $[a, b]$ à valeurs dans l'intervalle $[a, b]$. Soit $u_0 \in [a, b]$. On définit alors par récurrence la suite de terme général u_i en posant pour tout $i \geq 1$ $u_i = f(u_{i-1})$. On étudie la convergence de cette suite. Nous allons montrer que sous certaines conditions on peut affirmer qu'il existe un point fixe pour f, c'est-à-dire une solution de l'équation $f(x) = x$ (cf. figure 5.2). Donnons tout d'abord le théorème général suivant, qui met en avant la notion de fonction lipschitzienne dont nous avons déjà parlé :

Théorème 5.3 (Théorème du point fixe) *Soit f une application d'un segment $[a, b]$ dans lui-même. On suppose que pour tout $n \geq 1$ la fonction f_n obtenue en itérant n fois f :*

$$f_n = f \circ f \circ \cdots \circ f$$

est lipschitzienne de rapport k_n. On suppose en outre que la série $\sum_n k_n$ est convergente. Alors la fonction f admet un point fixe c et un seul. Pour tout

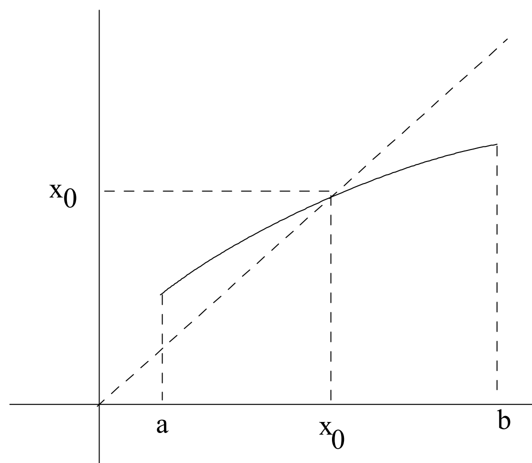


FIG. 5.2 – Théorème du point fixe

point $x_0 \in [a, b]$ la suite $(x_n)_{n\geq 0}$ définie par récurrence à partir de ce point x_0 en itérant f :

$$x_n = f(x_{n-1}) = f_n(x_0),$$

converge vers le point fixe c.

Preuve. Soient $m > n$ deux indices. On peut écrire successivement :

$$|x_m - x_n| \leq \sum_{i=n}^{m-1} |x_{i+1} - x_i|,$$

$$|x_m - x_n| \leq \sum_{i=n}^{m-1} k_i |x_1 - x_0|,$$

$$|x_m - x_n| \leq (b - a) \sum_{i=n}^{m-1} k_i.$$

Comme la série de terme général k_i est convergente, cette inégalité prouve que la suite $(x_n)_{n\geq 0}$ est une suite de Cauchy, donc converge vers un point c du segment $[a, b]$. La fonction f étant continue (puisque lipschitzienne), on conclut que $c = f(c)$. Ce point est l'unique point fixe. En effet soit d un point fixe, et soit n un indice tel que $k_n < 1$. alors $|f_n(c) - f_n(d)| \leq k_n|c - d|$, puisque f_n est k_n-lipschitzienne, et aussi $|f_n(c) - f_n(d)| = |c - d|$ puisque c et d sont des points fixes de f (et donc de f_n). On en déduit que :

$$|c - d| \leq k_n|c - d|,$$

avec $0 < k_n < 1$, donc $d = c$. ■

Un cas important d'application de de ce théorème est le cas où f est lipschitzienne de rapport $k < 1$. En effet dans ce cas, f_n est lipschizienne de rapport k_n avec $k_n \leq k^n$, et la série $\sum_n k_n$ est convergente.

Corollaire 5.1 *Soit f une application d'un segment $[a, b]$ dans lui-même. On suppose que la fonction f est lipschitzienne de rapport $k < 1$. Alors la fonction f admet un point fixe c et un seul. Pour tout point $x_0 \in [a, b]$ la suite $(x_n)_{n\geq 0}$ définie par récurrence à partir de ce point x_0 en itérant f :*

$$x_n = f(x_{n-1}),$$

converge vers le point fixe c.

En pratique, on montre souvent qu'une fonction est lipschitzienne de rapport k à l'aide du théorème des accroissements finis. On a donc le corollaire utile suivant :

Corollaire 5.2 *Soit f une application continue d'un segment $[a, b]$ dans lui-même. Supposons que f soit dérivable sur $]a, b[$ et de dérivée bornée par un nombre $k < 1$, c'est-à-dire :*

$$\sup_{x\in]a,b[} |f'(x)| = k < 1,$$

alors la fonction $f(x)$ admet un point fixe x_0 et un seul sur l'intervalle $[a, b]$. La suite $(u_i)_i$ définie précédemment converge vers le point fixe x_0.

Preuve. Le théorème des accroissements finis montre que f est lipschitzienne de rapport $k = \sup_{x\in]a,b[} |f'(x)|$. Le corollaire précédent permet de conclure. ■

5.4.2 Quelques exemples

Dans le premier exemple que nous allons donner, nos emploierons deux méthodes : nous ferons tout d'abord un calcul à la main, puis nous utiliserons directement le théorème précédent. Nous suggérons au lecteur de calquer ce shéma pour les autres exemples.

Exemple 1 : calcul du sinus et du cosinus de $1°$ par réinjection

▶ Présentation de l'exemple

Les propriétés géométriques des lignes trigonométriques permettent de calculer facilement $\sin(\pi/6)$, $\sin(\pi/4)$, $\sin(\pi/3)$, $\sin(\pi/5)$ ainsi évidemment que les

cosinus des mêmes angles. En utilisant la formule qui donne le sinus d'une différence on trouve $\sin(\pi/30)$, en utilisant la formule qui donne le sinus de l'arc moitié on obtient $\sin(\pi/60)$. On a donc une table donnant les sinus (et les cosinus) de 3° en 3°. Il faudrait donc calculer $\sin(\pi/180)$ pour avoir une table trigonométrique donnant les lignes de tous les angles en degré entier. On utilise alors la formule :

$$\sin(3x) = 3\sin(x) - 4\sin^3(x) \tag{5.2}$$

qu'on va appliquer ici avec $x = \pi/180$. On cherche alors à résoudre cette équation connaissant la valeur $a = \sin(3x)$. L'équation se ramène à :

$$\sin(x) = \frac{1}{3}(4\sin^3(x) + a).$$

Donc si on pose :

$$f(u) = \frac{1}{3}(4u^3 + a),$$

le nombre $\sin(\pi/180)$ est solution de l'équation :

$$f(u) = u,$$

c'est-à-dire est un point fixe de f.

▶ **Calcul à la main**

Pour calculer cette solution on part d'une valeur approchée de la solution u_0, par exemple $u_0 = 0$, et on calcule $u_1 = f(u_0)$. La valeur u_1 est réinjectée dans le second membre pour calculer une nouvelle approximation $u_2 = f(u_1)$, puis plus généralement, par récurrence :

$$u_n = f(u_{n-1}).$$

Théorème 5.4 *La suite de terme général u_n converge vers la valeur cherchée $\sin(\pi/180)$.*

Preuve. La fonction f a pour dérivée $f'(u) = 4u^2$, donc elle est croissante. On a par ailleurs :

$$u_1 = f(0) = \frac{a}{3} \geq 0,$$

et aussi :

$$f\left(\frac{a}{2}\right) = \frac{1}{3}\left(4\frac{a^3}{8} + a\right),$$

$$f\left(\frac{a}{2}\right) = \frac{a^3}{6} + \frac{a}{3},$$

$$f\left(\frac{a}{2}\right) \leq \frac{a}{6} + \frac{a}{3},$$

donc :

$$f\left(\frac{a}{2}\right) \leq \frac{a}{2}.$$

On conclut que l'image par f de l'intervalle $[0, a/2]$ est incluse dans l'intervalle $[0, a/2]$, autrement dit, restreinte à l'intervalle $[0, a/2]$, f est une application de $[0, a/2]$ dans lui-même et cette application est croissante.
Comme $u_0 = 0$, $u_1 = f(u_0) \geq u_0$, on a aussi $f(u_1) \geq f(u_0)$, c'est-à-dire que $u_2 \geq u_1$ et par récurrence $u_n \geq u_{n-1}$. La suite de terme général u_n est croissante, elle est majorée par $a/2$, elle converge donc vers une limite b. Cette limite vérifie $b = f(b)$. L'étude de la fonction $g(u) = u - f(u)$ sur l'intervalle $[0, a/2]$ nous montre que l'équation $g(u) = 0$ a au plus une solution dans cet intervalle (en effet $g(u)$ est strictement croissante sur cet intervalle). Donc le b trouvé comme limite de la suite de terme général u_n est l'unique solution de l'équation $u = f(u)$ sur l'intervalle $[0, a/2]$. Comme on sait que $\sin(\pi/180)$, qui est dans l'intervalle en question, est solution de cette équation, pas de doute, $b = \sin(\pi/180)$. ■

▶ **Application du théorème général**

Il suffit de voir que sur l'intervalle $[0, a/2]$ la dérivée de la fonction f vérifie :

$$\left|f'(x)\right| \leq a^2 < 1.$$

Le corollaire 5.2 permet de conclure.

▶ **Qualité de l'approximation**

On peut aussi évaluer la qualité de l'approximation, Sur l'intervalle $[0, a/2]$:

$$\left|f'(x)\right| \leq a^2.$$

On a alors en utilisant l'inégalité des accroissements finis et le fait que $b = f(b)$:

$$|b - u_n| = |f(b) - f(u_{n-1})| \leq a^2|b - u_{n-1}|,$$

et en réitérant le procédé,

$$|b - u_n| \leq (a^2)^n|b - u_0|,$$

et en majorant $|b - u_0|$ par la longueur de l'intervalle :

$$|b - u_n| \leq \frac{a^{2n+1}}{2}.$$

(Ceci revient à refaire la démonstration du théorème général.)

Exemple 2 : calcul de l'inverse d'un nombre.

▶ **Présentation de l'exemple**

Soit a un nombre réel que nous supposerons > 0 dont nous voulons calculer l'inverse $1/a$. La méthode que nous présentons, et qui est implémentée dans certaines calculatrices et certains langages de programmation, est très performante. C'est la méthode de Newton, développée pour le cas particulier qui nous préoccupe.
Introduisons la fonction :

$$F(x) = x(2 - ax).$$

Cherchons les solutions de l'équation :

$$x = F(x) \tag{5.3}$$

c'est-à-dire de :

$$x(1 - ax) = 0. \tag{5.4}$$

Les solutions sont 0 et $1/a$. Nous allons essayer d'approcher la solution $1/a$. L'idée est de partir d'une valeur approchée $u_0 > 0$ de $1/a$, de calculer le premier terme $u_1 = F(u_0)$, de réinjecter u_1 dans la partie droite de l'équation 5.3 pour obtenir $u_2 = F(u_1)$ et plus généralement par récurrence $u_n = F(u_{n-1})$.
Notons I l'intervalle ouvert $]0, 2/a[$.

Théorème 5.5 *Soit $u_0 \in I$. Alors la suite définie par le premier terme u_0 et la relation de récurrence pour $n \geq 1$*

$$u_n = F(u_{n-1}),$$

converge vers 1/a.

Preuve. La dérivée de $F(x)$ sur l'intervalle I est donnée par :

$$F'(x) = 2(1 - ax).$$

Sur l'intervalle I, la fonction est positive et atteint son maximum en $x = 1/a$. En ce point la valeur de la fonction est $1/a$. On peut donc conclure que l'image de l'intervalle I est incluse dans I. Remarquons aussi que sur l'intervalle $]0, 1/a]$ la fonction $F(x)$ est strictement croissante. Si $u_0 \in I$ alors $0 < F(u_0) \leq 1/a$. Posons ensuite $u_1 = F(u_0)$ et par récurrence $u_n = F(u_{n-1})$. Tous les termes $u_1, u_2, \cdots$ sont donc dans l'intervalle $]0, 1/a]$. De ce fait, $au_n \leq 1$, si bien que $(2 - au_n) \geq 1$ et donc

$$u_{n+1} = u_n(2 - au_n) \geq u_n.$$

La suite $(u_n)_{n\geq 1}$ est croissante, majorée par $1/a$, donc converge vers une limite $0 < b \leq 1/a$ qui vérifie $b = b(2 - ab)$.
Cette limite est donc le point $1/a$ de F. ■

On peut étudier aussi la qualité de l'approximation obtenue. Pour cela on suppose qu'on parte d'une valeur u_0 telle que

$$0 \leq u_0 \leq \frac{1}{a}.$$

Dans ces conditions on sait que toutes les valeurs u_n sont aussi dans cet intervalle. Alors on a successivement :

$$1 - au_n = 1 - au_{n-1}(2 - au_{n-1}),$$

$$1 - au_n = 1 - 2au_{n-1} + a^2u_{n-1}^2,$$

$$1 - au_n = (1 - au_{n-1})^2,$$

$$1 - au_n = (1 - au_0)^{2^n}.$$

On obtient donc :

$$\frac{1}{a} - u_n = \frac{1}{a}(1 - au_0)^{2^n}.$$

Posons $\alpha = 1 - au_0$. Compte tenu des conditions sur la valeur initiale u_0 le nombre α vérifie $0 \leq \alpha < 1$. La suite u_n qui vérifie :

$$\left|\frac{1}{a} - u_n\right| = \frac{1}{a}\alpha^{2^n}$$

a donc une convergence quadratique vers $1/a$.
Lors d'une implémentation effective, le problème du choix du point de départ se pose. On commence par écrire a sous la forme :

$$a = c\,10^k,$$

où $1 \leq c < 10$ et où k est un entier relatif. Ceci permet de se ramener à calculer l'inverse d'un nombre compris entre 1 et 10. Ceci étant fait, on peut supposer désormais que

$$1 \leq a < 10.$$

Choisissons u_0 de la façon suivante :

1. si $1 \leq a < 2$ alors $u_0 = 0.5$,
2. si $2 \leq a < 3$ alors $u_0 = 0.3$,
3. si $3 \leq a < 5$ alors $u_0 = 0.2$,

4. si $5 \leq a < 10$ alors $u_0 = 0.1$.

Ce choix respecte la condition $u_0 \leq 1/a$, et de plus $au_0 \geq 0.5$, par conséquent $1 - au_0 \leq 0.5$. Ces choix étant faits on obtient puisque $1/a \leq 1$ et $\alpha \leq 1/2$:

$$\left|\frac{1}{a} - u_n\right| \leq \left(\frac{1}{2}\right)^{2^n}.$$

5.4.3 La méthode de Newton

On peut se demander à propos de l'exemple précédent pourquoi la convergence est si rapide. C'est ce que nous allons essayer d'expliquer maintenant.
On peut déjà, compte tenu de la situation, penser que la convergence va être rapide. En effet compte tenu de l'inégalité des acroissements finis :

$$u_n - \frac{1}{a} = F(u_{n-1}) - F\left(\frac{1}{a}\right) \leq k\left|u_{n-1} - \frac{1}{a}\right|,$$

où k est un majorant de $F'(x)$ sur l'intervalle considéré. Donc :

$$\left|u_n - \frac{1}{a}\right| \leq k^n \left|u_0 - \frac{1}{a}\right|,$$

si bien que si $k < 1$ la suite converge. Mais ici, la dérivée est nulle au point fixe, donc, on peut s'attendre à une amélioration de la rapidité d'approximation lorsqu'on se rapproche du point fixe. On a vu en fait que l'approximation est quadratique.
Généralisons cette méthode. Soit $f(x)$ une fonction sur un intervalle $I = [a, b]$ et ayant sur cet intervalle un seul zéro c. On suppose que $f'(x)$ ne s'annule pas sur I.

$$F(x) = x - \frac{f(x)}{f'(x)}.$$

Alors c est le seul point fixe de $F(x)$ sur l'intervalle I.

$$F'(x) = \frac{f(x)f''(x)}{f'^2}.$$

Donc $F'(c) = 0$. On se retrouve dans la situation favorable précédente.
Remarquons que cette construction a une interprétation géométrique simple : Il s'agit ici de remplacer la fonction f dont on cherche un zéro par sa tangente en un point voisin du zéro cherché (cf. figure 5.4).

Ainsi, à partir d'un point u_0 proche de la solution x_0 de l'équation

$$f(x) = 0$$

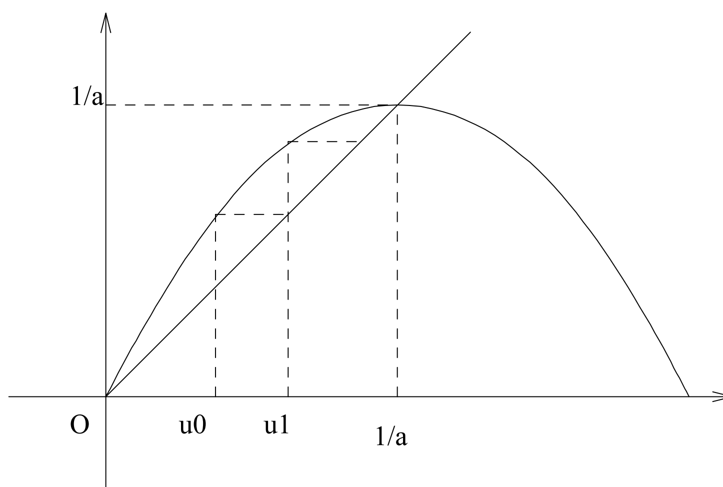


FIG. 5.3 – Approximation de $1/a$

(qu'on supposera unique, tout au moins dans un intervalle adapté), on construit la tangente à la courbe $y = f(x)$ au point $(u_0, f(u_0))$. Cette tangente coupe l'axe des abscisses au point u_1. On itère cette construction à partir de la valeur u_1 pour obtenir le point u_2. Le calcul de l'équation de la tangente au point d'abscisse x et de son intersection avec l'axe des x montre que si on introduit :

$$F(x) = x - \frac{f(x)}{f'(x)},$$

alors on peut écrire :

$$u_1 = F(u_0), \quad u_2 = F(u_1), \ldots \quad , u_n = F(u_{n-1}), \ldots$$

On vérifie imédiatement que x_0 est point fixe de la fonction $F(x)$. De plus si on suppose la fonction f de classe C^2 par exemple et si on suppose que la dérivée f' de f ne s'annule pas, alors $F(x)$ est dérivable et sa dérivée est nulle au point x_0. Il y aura donc un voisinage fermé de ce point fixe, pas nécessairement simple à déterminer effectivement, dans lequel la théorie du paragraphe précédent s'applique. De plus, comme la dérivée de F sera proche de zéro, on peut espérer une bonne convergence de la suite $(u_n)_n$, d'autant plus rapide qu'on va se rapprocher du point fixe.

$$F(x) - F(c) = F(x) - c = \frac{A(x)}{f'(x)}.$$

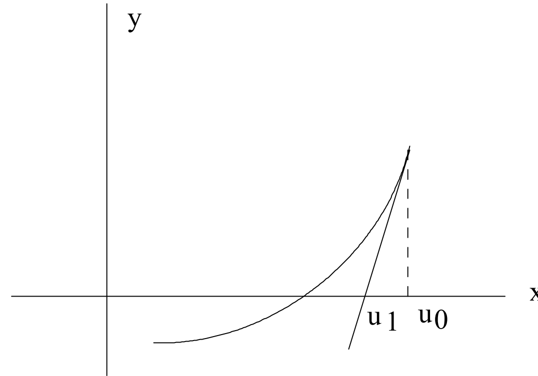


FIG. 5.4 – Méthode de Newton

Comme $F(c) - c = 0$ on conclut que $A(c) = 0$. Comme $F'(c) = 0$ on conclut que $A'(c) = 0$. Si par exemple $A(x)$ est un polynôme ça veut dire qu'il s'écrit :

$$A(x) = (x - a)^2 B(x),$$

où $B(x)$ est un polynôme et donc la convergence est quadratique. Nous verrons (application de la formule de Taylor) que ceci a lieu pour des fonctions plus générales.

Exemple 3 : Considérons la fonction $f(x) = x^2 - 2$. En nous restreignant à $x \geq 0$, nous allons déterminer la solution positive de l'équation $x^2 = 2$, c'est-à-dire $x = \sqrt{2}$. Introduisons donc, comme nous l'indique la méthode de Newton, la fonction :

$$F(x) = x - \frac{x^2 - 2}{2x} = x - \frac{x}{2} + \frac{1}{x} = \frac{1}{2}\left(x + \frac{2}{x}\right).$$

On voit que par exemple $F([1,2]) \subset [1,2]$ et sur cet intervalle la dérivée de $F(x)$ est en valeur absolue majorée par $1/2$. Donc le théorème du point fixe s'applique sur cet intervalle. Nous avons successivement :

$$\frac{1}{2}\left(u_n + \frac{2}{u_n}\right) - \sqrt{2} = F(u_n) - F(\sqrt{2}) = \frac{1}{2u_n}\left(u_n^2 + (\sqrt{2})^2 - 2u_n\sqrt{2}\right),$$

$$\frac{1}{2}\left(u_n + \frac{2}{u_n}\right) - \sqrt{2} = \frac{1}{2u_n}\left(u_n - \sqrt{2}\right)^2,$$

$$\left|\frac{1}{2}\left(u_n + \frac{2}{u_n}\right) - \sqrt{2}\right| \leq \frac{1}{2}\left(u_n - \sqrt{2}\right)^2,$$

c'est-à-dire :

$$|u_{n+1} - \sqrt{2}| \leq \frac{1}{2}\left(u_n - \sqrt{2}\right)^2.$$

La convergence est donc **quadratique**.

Exemple 4 : Prenons $f(x) = x^2 - x - 1$ et regardons ce qu'il se passe sur l'intervalle $[1, 2]$. On est alors amené à introduire :

$$F(x) = \frac{x^2 + 1}{2x - 1}.$$

Exemple 5 : On peut reprendre le calcul de $b = \sin(\pi/180)$ connaissant $a = \sin(\pi/60)$. Nous avions utilisé la fonction $f = 1/3(4x^3 + a)$ pour définir par récurrence la suite $u_n = f(u_{n-1})$ qui convergeait vers le point fixe $b = f(b)$. Hélas, au point b, la valeur de la dérivée $f'(b)$ n'est pas nulle, si bien que la convergence n'est peut être pas aussi rapide qu'elle pourrait l'être avec la méthode de Newton. Pour appliquer cette méthode, nous remarquons que nous cherchons le zéro de la fonction $h(x) = f(x) - x$. Conformément aux calculs précédent, nous introduisons la fonction :

$$g(x) = x - \frac{h(x)}{h'(x)}.$$

Tous calculs faits, on obtient :

$$g(x) = \frac{8x^3 - a}{12x^2 - 3}.$$

La suite $u_0 = 0$, $u_n = g(u_{n-1})$ converge alors plus vite que précédemment vers la valeur cherchée $b = \sin(\pi/180)$. Une expérimentation avec un système de calcul confirme ce comportement. En partant de $u_0 = 0$, la première itération, $u_1 = g(u_0)$, donne 4 décimales exactes, la deuxième, $u_2 = g(u_1)$, donne 11 décimales exactes.

5.5 Intégration, outils de base

L'intégrale (quand elle existe), est un outil régularisant. Il effectue une moyenne et gomme un certain nombre d'irrégularités de la fonction qu'on intègre. Prenons par exemple la fonction en escalier définie sur $[0, 1]$ par :

$$f(x) = \begin{cases} 0 & \text{si } 0 \leq x < 1/2 \\ 1 & \text{si } 1/2 \leq x \leq 1. \end{cases}$$

Cette fonction f a un point de discontinuité en 1/2 et peut être intégrée :

$$g(x) = \int_0^x f(t)dt = \begin{cases} 0 & \text{si } 0 \leq x < 1/2 \\ x - 1/2 & \text{si } 1/2 \leq x \leq 1. \end{cases}$$

La fonction $g(x)$ est continue sur $[0, 1]$. Mais $g(x)$ n'est pas dérivable (au point 1/2).
Remarquons que la fonction f n'a pas de primitive sur $[a, b]$. Ce n'est pas une dérivée. D'ailleurs on peut montrer qu'une dérivée vérifie le théorème des valeurs intermédiaires[2], ce qui n'est pas le cas de f ici :

Théorème 5.6 *Soit f une fonction définie sur un intervalle (a, b) qui est une dérivée. Soient α et β deux points de (a, b). Alors $f(x)$ prend toute valeur comprise entre $f(\alpha)$ et $f(\beta)$.*

5.5.1 Un problème de raccord

On considère sur l'intervalle $[-1, 1]$ la fonction paire f définie sur $[0, 1]$ par $f(x) = 1 - x$. On considère les points P et Q de coordonnées respectives $(-h, f(h))$, $(h, f(h))$ où $0 < h < 1$. On cherche un arc de parabole qui se raccorde en ces deux points en étant tangent aux deux droites constituant la courbe représentative de f. La dérivée de la fonction parabolique sur $[-h, h]$ est donc $g'(x) = -x/h$ (polynôme du premier degré qui vaut 1 en $-h$ et -1 en h). Donc $g(x) = -x^2/2h + C$. La constante est donnée par $g(h) = f(h) = 1 - h$, ce qui donne $C = 1 - h/2$. En définitive :

$$g(x) = -\frac{x^2}{2h} + 1 - \frac{h}{2}.$$

Bien entendu on aurait pu directement écrire les conditions sur g (g polynôme du second degré, $g'(-h) = 1$, $g'(h) = -1$), mais cette méthode très simple qui consiste à déterminer g à partir de sa dérivée est intéressante.

5.5.2 Intégration des relations de comparaison

L'intégration se comporte bien vis à vis des opérations de comparaisons classiques :

▶ **Inégalités**

Théorème 5.7 *Si f et g sont des fonctions intégrables sur $[a, b]$ et si $f \leq g$, alors $\int_a^b f(t)dt \leq \int_a^b g(t)dt$. Ce résultat persiste pour les intégrales généralisées.*

[2] Il s'agit du lemme de Darboux (voir lemme 16 de "*L'épreuve d'exposé au CAPES mathématiques, vol. IV*", de D.-J. Mercier).

▶ **Comportement asymptotique : cas de la divergence**

Regardons maintenant ce qui se passe du point de vue des comparaisons asymptotiques des intégrales généralisées. On considèrera qu'on regarde le comportement au voisinage de $+\infty$ (mais ce pourrait être au voisinage d'un point a). On utilise les notations de Landau $o(g)$ et $O(g)$ ainsi que l'équivalence $\sim$.

Théorème 5.8 *Soit g une fonction continue > 0 sur $[a, +\infty[$ telle que*

$$\int_a^{+\infty} g(t)dt = +\infty.$$

Soit f une fonction continue sur $[a, +\infty[$.

(1) Si $f = O(g)$ alors

$$\int_a^x f(t)dt = O\left(\int_a^x g(t)dt\right)$$

(2) Si $f = o(g)$ alors

$$\int_a^x f(t)dt = o\left(\int_a^x g(t)dt\right)$$

(3) Si $f \sim g$ alors

$$\int_a^x f(t)dt \sim \int_a^x g(t)dt.$$

▶ **Comportement asymptotique : cas de la convergence**

Théorème 5.9 *Soit g une fonction continue > 0 sur $[a, +\infty[$ telle que*

$$\int_a^{+\infty} g(t)dt < +\infty.$$

Soit f une fonction continue sur $[a, +\infty[$.

(1) Si $f = O(g)$ alors

$$\int_x^{+\infty} f(t)dt = O\left(\int_x^{+\infty} g(t)dt\right)$$

(2) Si $f = o(g)$ alors

$$\int_x^{+\infty} f(t)dt = o\left(\int_x^{+\infty} g(t)dt\right)$$

(3) Si $f \sim g$ alors

$$\int_x^{+\infty} f(t)dt \sim \int_x^{+\infty} g(t)dt.$$

Ces divers résultats permettent, en collaboration avec l'intégration par partie de déterminer des comportements asymptotiques pour diverses intégrales. On verra des exemples dans la section consacrée à l'intégration par partie.

5.5.3 Calcul du sinus et du cosinus de $1°$ par une approximation polynomiale

On va prendre ici pour valeur approchée du sinus au voisinage de 0 le polynôme :

$$P(x) = x - \frac{x^3}{6}.$$

Compte tenu du problème posé on prendra $x \geq 0$.

Pour évaluer la qualité de cette approximation nous allons partir de l'inégalité triviale

$$\cos(x) \leq 1.$$

Par conservation des inégalités par intégration sur un intervalle $[a, b]$ où $b \geq a$, on a donc :

$$\int_0^x \cos(t)\, dt \leq \int_0^x dt,$$

c'est-à-dire :

$$\sin(x) \leq x.$$

En réappliquant cette méthode on obtient maintenant :

$$\int_0^x \sin(t)\, dt \leq \int_0^x x\, dt,$$

ou encore :

$$1 - \cos(x) \leq \frac{x^2}{2},$$

et en conséquence l'encadrement suivant :

$$1 - \frac{x^2}{2} \leq \cos(x) \leq 1.$$

On obtient alors successivement, toujours par intégration les encadrements suivants :

$$x - \frac{x^3}{6} \leq \sin(x) \leq x,$$

$$1-\frac{x^2}{2}\le\cos(x)\le 1-\frac{x^2}{2}+\frac{x^4}{24},$$
$$x-\frac{x^3}{6}\le\sin(x)\le x-\frac{x^3}{6}+\frac{x^5}{120}.$$

Ceci montre en particulier que :

$$0\le\sin(x)-P(x)\le\frac{x^5}{120}.$$

Appliqué au cas où $x=\pi/180$, on obtient :

$$0\le\sin\left(\frac{\pi}{180}\right)-P\left(\frac{\pi}{180}\right)\le 1.35\,10^{-11}.$$

5.6 Quelques classes habituelles de fonctions

Quand on fait de l'analyse on essaie le plus possible de "rendre cohérents" les systèmes avec lesquels on travaille. Précisons un peu ce que j'entend par là. On est souvent confronté au problème suivant : on a un ensemble d'objets mathématiques (par exemple des fonctions continues sur un segment $[a,b]$). On a un outil qui s'applique à ces objets (par exemple l'intégrale sur le segment $[a,b]$). Enfin on a un mode de convergence (par exemple la convergence uniforme sur $[a,b]$). La question est : l'outil agit il de manière stable vis à vis de la convergence dans cet ensemble ? (Par exemple : soit f_n une suite de fonctions continues sur $[a,b]$ qui converge uniformément vers f. La suite des intégrales des f_n converge-t-elle vers l'intégrale de f ?). Dans certains cas ceci n'a pas lieu, et en fait les raisons du dysfonctionnement peuvent être multiples. Prenons l'exemple de l'espace des fonctions continues sur un segment $[a,b]$, et la convergence simple.

1. Le premier obstacle est que cet espace n'est pas stable pour la convergence simple.

Exemple 1 : on considère la suite de fonctions $(f_n)_{n>0}$ définies sur $[0,1]$ par :

$$f_n(x)=\begin{cases}-nx+1 & \text{si } x\in\left[0,\frac{1}{n}\right]\\ 0 & \text{sinon .}\end{cases}$$

Cette suite de fonctions continues converge simplement vers la fonction discontinue :

$$f(x)=\begin{cases}1 & \text{si } x=0\\ 0 & \text{sinon .}\end{cases}$$

2. On peut alors penser à considérer l'espace des fonctions intégrales au sens de Riemann sur $[a,b]$. Hélas, il n'est pas stable non plus pour la convergence

simple : On peut trouver une suite de fonctions intégrables au sens de Riemann, qui converge simplement vers une fonction bornée, qui n'est pas intégrable au sens de Riemann. Pour cela considérons la suite de fonctions f_n, définies sur $[0,1]$ par :

$$f_n(x) = \begin{cases} 1 & \text{si } x = \dfrac{k}{n!} \text{ où } 0 \leq k \leq n! \\ 0 & \text{sinon .} \end{cases}$$

Cette suite converge simplement vers la fonction de Dirichlet :

$$f(x) = \begin{cases} 1 & \text{si } x \in [0,1] \cap \mathbb{Q} \\ 0 & \text{sinon .} \end{cases}$$

Les fonctions f_n sont intégrables (au sens de Riemann), la fonction f ne l'est pas.
D'autre part, rester dans le cadre des fonctions continues et de la convergence uniforme est un peu contraignant. Le cadre des fonctions réglées est un bon compromis.

Définition 5.2 *Soit f une fonction définie sur un segment $[a,b]$. La fonction f est dite réglée, si en tout point $x \in]a,b[$, elle admet une limite à droite et une limite à gauche, et si elle admet une limite à droite en a et une limite à gauche en b.*

On doit donner tout de suite une caractérisation très importante des fonctions réglées, qui pourrait d'ailleurs servir de définition.

Théorème 5.10 *Une fonction f définie sur un segment $[a,b]$ est réglée, si et seulement si elle est limite uniforme de fonctions en escalier.*

Cette caractérisation nous fait comprendre l'intérêt de cette classe lorsqu'on travaille avec l'intégrale de Riemann, qui rappelons le, est définie par passage à la limite sur des intégrales de fonctions en escalier.
Les fonctions en escaliers, les fonctions continues, les fonctions monotones sont des fonctions réglées. Cette classe est stable pour la convergence uniforme et l'intégrale d'une limite de fonctions réglées est la limite des intégrales.

5.7 L'intégration par partie

Nous présentons ici un outil très simple mais particulièrement efficace : l'intégration par partie. À son propos, je citerais Roger Godement qui dans la

rubrique *Calcul Infinitésimal* de l'Encyclopédie Universalis écrit : "[l'intégration par partie est] très commode pour le calcul pratique des intégrales, mais dont l'intérêt est ailleurs lorsqu'on s'occupe de mathématiques".
Donc, si l'on veut bien laisser de côté le volet anecdotique du calcul exact de certaines primitives, il faut comprendre l'intégration par partie comme l'écriture d'une expression sous forme d'une partie principale (la partie tout intégrée) et d'un reste (la deuxième intégrale)

$$\int_a^x f'(t)g(t)dt = [f(t)g(t)]_a^x - \int_a^x f(t)g'(t)dt.$$

Évidemment, si on espère que la partie

$$-\int_a^x f(t)g'(t)dt$$

puisse être considérée comme un reste, il faut qu'elle soit négligeable devant la partie tout intégrée

$$[f(t)g(t)]_a^x.$$

Cette remarque nous indique comment choisir, quand on fait une intégration par partie d'un produit de deux fonctions, celle qu'on intègre et celle qu'on dérive : celle qu'on dérive est en général celle qui relativement varie peu, de manière à obtenir une dérivée petite devant la fonction elle-même. Attention ceci ne s'applique pas pour l'utilisation de l'intégration par partie pour des calculs de primitives, ni pour l'établissement de formules théoriques où l'on veut obtenir une forme particulière pour le terme tout intégré.
Dans la suite de ce chapitre, nous donnons deux applications très importantes de l'intégration par partie : la formule de Taylor et la formule d'Euler-Maclaurin. Ces deux formules constituent des outils de base des méthodes d'approximation.

5.7.1 Intégrons dans le bon sens

Exemple 1. Il s'agit d'évaluer au voisinage de $+\infty$ l'intégrale

$$\int_{\ln(x)}^{+\infty} \frac{e^{-u}}{u^2} du.$$

On choisit clairement de dériver la fonction $1/u^2$ et d'intégrer la fonction e^{-u}, on obtient

$$\int_{\ln(x)}^{+\infty} \frac{e^{-u}}{u^2} du = \frac{1}{x(\ln(x))^2} - 2\int_{\ln(x)}^{+\infty} \frac{e^{-u}}{u^3} du.$$

Cette manière de faire permet effectivement d'obtenir un reste négligeable devant la partie intégrée. En effet :

$$\frac{e^{-u}}{u^3} = o\left(\frac{e^{-u}}{u^2}\right),$$

donc :

$$\int_{\ln(x)}^{+\infty} \frac{e^{-u}}{u^3} du = o\left(\int_{\ln(x)}^{+\infty} \frac{e^{-u}}{u^2} du\right)$$

et en conséquence :

$$\int_{\ln(x)}^{+\infty} \frac{e^{-u}}{u^2} du \sim \frac{1}{x(\ln(x))^2}.$$

Exemple 2. Dans cet exemple nous cherchons la nature de l'intégrale :

$$\int_1^{+\infty} \frac{sin(x)}{x} dx.$$

On étudie donc :

$$\lim_{A \to +\infty} \int_1^A \frac{sin(x)}{x} dx.$$

On va faire une intégration par partie, en dérivant la fonction qui varie peu au voisinage de $+\infty$ et en intégrant l'autre, c'est-à-dire $sin(x)$. On a donc en détaillant un peu l'intégration par partie sur l'intervalle $[0, A]$:

$$\begin{array}{ll} u = \frac{1}{x} & u' = -\frac{1}{x^2} \\ v = -cos(x) & v' = sin(x) \end{array}$$

$$\int_1^A \frac{sin(x)}{x} dx = \left[-\frac{cos(x)}{x}\right]_1^A - \int_1^A \frac{cos(x)}{x^2} dx.$$

La partie principale tend vers $cos(1)$ lorsqu'on fait tendre A vers $+\infty$, le reste, c'est-à-dire :

$$-\int_1^A \frac{cos(x)}{x^2} dx$$

est majoré en valeur absolue par une intégrale convergente :

$$\left|\int_1^A \frac{cos(x)}{x^2} dx\right| \leq \int_1^A \left|\frac{cos(x)}{x^2}\right| dx \leq \int_1^{+\infty} \frac{1}{x^2} dx,$$

donc constitue une intégrale absolument convergente sur $[1, +\infty[$.
On en conclut que l'intégrale étudiée est convergente.

On peut montrer que l'intégrale étudiée n'est pas absolument convergente en étudiant les sommes partielles :

$$S_n = \int_1^{n\pi} \frac{|sin(x)|}{x} dx,$$

$$S_n = \int_1^{\pi} \frac{|sin(x)|}{x} dx + \sum_{k=1}^{n-1} \int_{k\pi}^{(k+1)\pi} \frac{|sin(x)|}{x} dx. \tag{5.5}$$

Or :

$$\int_{k\pi}^{(k+1)\pi} \frac{|sin(x)|}{x} dx \geq \frac{1}{(k+1)\pi} \int_0^{\pi} sin(x) dx = \frac{2}{(k+1)\pi},$$

ce qui prouve que la série qui intervient dans la formule (5.5) est divergente.

Exemple 3. Dans cet exemple nous cherchons un équivalent au voisinage de $+\infty$ de

$$\int_a^x \frac{dt}{\ln(t)}.$$

(où $a > 1$).
On effectue une intégration par partie, où on dérive $1/\ln(t)$ et on intègre 1 :

$$\int_a^x \frac{dt}{\ln(t)} = \left[\frac{t}{\ln(t)} \right]_a^x + \int_a^x \frac{dt}{(\ln(t))^2}.$$

Comme :

$$\frac{1}{(\ln(t))^2} = o\left(\frac{1}{\ln(t)} \right),$$

on a :

$$\int_a^x \frac{dt}{(\ln(t))^2} = o\left(\int_a^x \frac{dt}{\ln(t)} \right).$$

En conséquence :

$$\int_a^x \frac{dt}{\ln(t)} \sim \frac{x}{\ln(x)}.$$

Exemple 4 (l'art de se faire enfumer). Il s'agit de calculer

$$I_p = \int_1^e x^2 (\ln(x))^p dx.$$

On a $I_0 = (e^3 - 1)/3$.

▶ **Le piège**

Le piège consiste ici à voir apparaître une formule très simple et à se laisser aller à intégrer par partie "à l'envers". C'est-à-dire qu'on va dériver $(\ln(x))^p$ et intégrer x^2.

$$I_p = \frac{e^3}{3} - \frac{p}{3} I_{p-1}.$$

Si nous itérons le calcul nous sommes amenés à écrire la formule exacte

$$I_p = \frac{e^3}{3}\left(1 - \frac{p}{3} + \frac{p(p-1)}{9} + \cdots + (-1)^{p-1}\frac{p!}{3^{p-1}}\right) + (-1)^p \frac{p!}{3^{p-1}}(e^3 - 1).$$

Hélas, comme à chaque étape on a pris un "reste plus grand que la partie qui aurait dû être principale", cette formule est très instable numériquement. En effet si on change un peu la valeur initiale I_0 en J_0, alors le terme J_p calculé vérifie

$$J_p - I_p = (-1)^p \frac{p!}{3^p}(J_0 - I_0),$$

ce qui fait que si $J_0 \neq I_0$, alors $|J_p - I_p| \to +\infty$. Donc une erreur d'arrondi va complètement modifier le calcul.

▶ **L'intégration dans le bon sens**

Bien sûr, si on intègre dans le bon sens en exhibant une partie principale et un reste plus petit que cette partie principale, alors ceci ne se produit plus. Intégrons donc $\frac{(\ln(x))^p}{x}$ et dérivons x^3. Nous obtenons

$$I_p = \frac{e^3}{p+1} - \frac{3}{p+1} I_{p+1},$$

et cette fois-ci en itérant le calcul on tombe sur une formule stable qui nous permet d'écrire tout de suite

$$I_p \sim \frac{e^3}{p}.$$

Si on veut pousser le développement plus loin on obtient :

$$I_p = e^3\left(\frac{1}{p} - \frac{4}{p^2}\right) + o\left(\frac{1}{p^2}\right).$$

Bien sûr la formule de récurrence trouvée est la même que dans le premier calcul, écrite différemment. Mais justement ceci nous permet de voir que si on a une vision "à l'envers" de l'intégration par partie, alors la suite des calculs qu'on est amené naturellement à faire conduit à de mauvaises situations.

5.7.2 Conclusion et remarques

Nous avons vu que l'intégration par partie est un outil puissant pour étudier des comportements asymptotiques d'intégrales convergentes ou divergentes en considérant la partie tout intégrée comme la partie principale et la deuxième intégrale comme un reste.
D'un autre côté, en analyse il y a un aller-retour pemanent entre des situations « discrètes » et des situations « continues ». On approche des fonctions en regardant ses valeurs en un nombre fini de points par exemple. Ou bien on construit une fonction en interpolant à partir d'un nombre fini de valeurs. On associe à une équation différentielle $y' = f(y)$ une suite telle que $u_0 = y_0$ et $u_{n+1} - u_n = hf(u_n)$, remplaçant ainsi la dérivée par une différence finie (c'est la méthode de la tangente d'Euler). En ce qui concerne les intégrales généralisées, elles trouvent leur équivalent discret en considérant les séries. On peut donc se demander quel est le pendant sur les séries de l'intégration par partie. C'est la transformation d'Abel. Considérons la série de terme général $a_n b_n$. Posons alors $S_k = \sum_{i=1}^{k} b_i$ et $S_0 = 0$. Alors :

$$\sum_{k=1}^{n} a_k b_k = \sum_{k=1}^{n} a_k (S_k - S_{k-1}),$$

$$\sum_{k=1}^{n} a_k b_k = \sum_{k=1}^{n} a_k S_k - \sum_{k=1}^{n} a_k S_{k-1} = \sum_{k=1}^{n} a_k S_k - \sum_{k=0}^{n-1} a_{k+1} S_k,$$

$$\sum_{k=1}^{n} a_k b_k = a_n S_n + \sum_{k=1}^{n-1} S_k (a_k - a_{k+1}).$$

On peut aussi écrire cette transformation pour un reste partiel :

$$\sum_{k=m}^{n} a_k b_k = a_n S_n - a_{m-1} S_{m-1} + \sum_{k=m}^{n-1} S_k (a_k - a_{k+1}).$$

À partir de là et de quelques hypothèses sur les comportements de a_n et S_n, on peut aussi énoncer des résultats sur le comportement asymptotique de la série de terme général $a_n b_n$.

5.8 La formule de Taylor

5.8.1 Remarque préliminaire

Comme nous allons le voir, la formule de Taylor peut être considérée comme une généralisation du théorème fondamental du calcul différentiel et intégral

qui permet d'écrire lorsque f est dérivable et de dérivée intégrable :

$$f(x) - f(a) = \int_a^x f'(t)dt.$$

En intégrant par partie l'intégrale du second membre, puis les intégrales calculées successivement en itérant le procédé, on obtient la formule de Taylor, qui nous permet d'exprimer une fonction comme un polynôme plus un reste.

5.8.2 Le théorème principal

Théorème 5.11 *Soit f une fonction à valeurs réelles définie sur le segment $[a-\epsilon, a+\epsilon]$ et $n+1$ fois continument dérivable sur ce segment. Alors pour tout point x du segment $[a-\epsilon, a+\epsilon]$ on peut écrire*

$$f(x) = f(a) + \frac{(x-a)}{1!}f(a) + \cdots + \frac{(x-a)^n}{n!}f^{(n)}(a) + \int_a^x \frac{(x-t)^n}{n!}f^{(n+1)}(t)dt.$$

Preuve. On utilise une démonstration par récurrence. La formule

$$\int_a^x f'(t)dt = f(x) - f(a),$$

assure que le théorème est vrai pour $n = 0$. Si on suppose vraie la formule à l'ordre $n \geq 0$ alors

$$\int_a^x \frac{(x-t)^n}{n!}f^{(n+1)}(t)dt = \left[\frac{-(x-t)^{n+1}}{(n+1)!}f^{(n+1)}(t)\right]_a^x + \int_a^x \frac{(x-t)^{n+1}}{(n+1)!}f^{(n+2)}(t)dt,$$

$$\int_a^x \frac{(x-t)^n}{n!}f^{(n+1)}(t)dt = \frac{(x-a)^{n+1}}{(n+1)!}f^{(n+1)}(a) + \int_a^x \frac{(x-t)^{n+1}}{(n+1)!}f^{(n+2)}(t)dt,$$

ce qui nous donne la formule à l'ordre $n+1$. ■

5.8.3 Obtention d'autres écritures

J'essaie autant que faire se peut d'exprimer les formules asymptotiques en utilisant des restes intégraux qui donnent des **formules exactes** et qui conduisent assez facilement à des majorations effectives. J'évite au maximum les formules faisant intervenir des restes écrits avec des points dont on ne connait pas la valeur mais juste une localisation. En général, dans un vrai problème on n'a jamais vraiment besoin de ces formules et les formules avec reste intégral ainsi que l'outil "intégration par partie" permettent d'obtenir les évaluations dont on a besoin. De toutes façons, il ne faut pas oublier que tant qu'on écrit la formule avec reste intégral **on ne perd rien**, toute l'information est là et reste

disponible sous une forme commode. Ce n'est pas le cas avec les autres écritures : formule de Taylor-Lagrange, formule de Taylor-Young, qui toutefois, sont difficiles à ommettre dans un cours ou dans une leçon formelle en raison de leur poids historique, mais que je ne conseille guère d'utiliser pour majorer effectivement. Insistons encore une fois sur le fait que lorsqu'on dispose du reste intégral (ou d'un autre type de reste d'ailleurs), la formule n'est utile que si on est capable de majorer ce reste et d'obtenir une bonne approximation polynomiale de la fonction f au voisinage du point a.

▶ **Formule de Taylor-Lagrange**

Cette formule est dans la ligne directe du théorème des accroissements finis dans sa version donnée par le théorème 5.1.

Théorème 5.12 *Soit f une fonction de classe C^n sur $[a,b]$ et admettant une dérivée d'ordre $n+1$ dérivable sur $]a,b[$. Alors, il existe un point $c \in]a,b[$ tel que :*

$$f(b) = f(a) + \frac{(b-a)}{1!} f(a) + \cdots + \frac{(b-a)^n}{n!} f^{(n)}(a) + \frac{(b-a)^{n+1}}{(n+1)!} f^{(n+1)}(c).$$

Cette formule se démontre en restant dans le cadre du calcul différentiel, en appliquant la formule des accroissements finis dans sa version donnée par le théorème 5.1. C'est une formule **globale**, qui permet d'avoir une évaluation sur tout un intervalle. Remarquons que l'interprétation géométrique qui faisait le grand intérêt du théorème des accroissements finis 5.1 a disparu ici. Dans la plupart des cas on gagnera à appliquer la formule avec reste intégral.

▶ **Formule de Taylor-Young**

Cette formule est une formule **locale** qui permet non pas d'avoir une évaluation sur tout un intervalle fixé à l'avance, mais un comportement au voisinage d'un point. Elle est destinée à écrire des développements limités et remplit parfaitement sa fonction dans ce cas en expédiant tout de suite le reste sous la forme voulue.

Théorème 5.13 *Soit f une fonction de classe C^n sur un intervalle I, et a un point de I. On suppose que f admet une dérivée d'ordre $n+1$ au point a. Alors pour tout $x \in I$ on a*

$$f(x) = f(a) + \frac{(x-a)}{1!} f(a) + \cdots + \frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}(a) + o\left(|x-a|^{n+1}\right).$$

L'étude locale d'une fonction, consistant le plus souvent en un développement limité à un ordre intéressant de la fonction au voisinage d'un point, est l'une des deux applications principales de la formule de Taylor, l'autre étant le développement en série entière.

5.8.4 Fonction développable en série entière

Le développement en série entière de fonctions est, comme nous l'avons souligné, l'une des applications principales de la formules de Taylor. On écrit, quand le développement est possible, l'égalité d'une fonction et de la somme d'une série entière sur tout un intervalle. Le seul espoir d'obtenir pour une fonction f un développement en série entière sur un intervalle ouvert I, est que f soit indéfiniment dérivable sur I, et dans ce cas, le développement de f autour d'un point a de l'intervalle I sera donné par :

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n. \tag{5.6}$$

À partir de là, plusieurs cas sont possibles :

(1) La série entière (5.6) a un rayon de convergence nul et donc f n'est pas développable en série entière au voisinage de a.

Exemple 1 : La série de fonctions de terme général $u_n(x) = e^{-n} sin(n^2 x)$ converge sur $\mathbb{R}$ vers une fonction f indéfiniment dérivable. La série de Taylor au point 0 de cette fonction a un rayon de convergence nul.

Exemple 2 : Le théorème de Borel affirme que si on se donne une série entière formelle, il existe une fonction f indéfiniment dérivable dont la série de Taylor en 0 soit la série entière fixée. À partir de là il est facile de montrer l'existence d'une fonction f dont la série de Taylor en 0 est $\sum_{n=0}^{\infty} n!x^n$, qui est de rayon de convergence nul.

(2) La série entière (5.6) a un rayon de convergence $R > 0$, mais quel que soit le voisinage V de a inclus dans le disque de convergence, f ne coïncide pas sur V avec la somme de la série entière. Dans ce cas, f n'est pas développable non plus au voisinage de a.

Exemple : la fonction définie par :

$$f(x) = \begin{cases} e^{-1/x^2} & \text{si} \quad x \neq 0 \\ 0 & \text{si} \quad x = 0 \end{cases}$$

est indéfiniment dérivable et toutes ses dérivées sont nulles en 0. La série de Taylor de f converge donc vers la fonction nulle sur tout intervalle, qui n'est en aucun cas f.

(3) La série entière (5.6) a un rayon de convergence $R > 0$ et il existe un voisinage V de a inclus dans le disque de convergence où f est la somme de la série entière. Dans ce cas f est développable en série entière au voisinage de a.

L'étude passe par l'observation du reste dans la formule de Taylor :

$$f(x) = f(a) + \frac{(x-a)}{1!} f(a) + \cdots + \frac{(x-a)^n}{n!} f^{(n)}(a) + \int_a^x \frac{(x-t)^n}{n!} f^{(n+1)}(t) dt.$$

En effet, si sur un intervalle $]a-h, a+h[$ on montre que le reste

$$R_n(x) = \int_a^x \frac{(x-t)^n}{n!} f^{(n+1)}(t)\, dt$$

converge vers 0, alors f est développable en série entière sur $]a-h, a+h[$.

5.8.5 Cas des dérivées positives

Le résultat suivant est assez intéressant et mérite d'être signalé.

Théorème 5.14 *Si f est indéfiniment dérivable sur $]-a, a[$ et a toutes ses dérivées positives sur $[0, a[$, alors f est somme de sa série de Taylor sur $[0, a[$.*

Preuve. Notons P_n le polynôme de Taylor de f en 0 à l'ordre n, et R_n le reste.

$$f(x) = P_n(x) + R_n(x),$$

avec

$$R_n(x) = \int_0^x \frac{(x-t)^n}{n!} f^{n+1}(t)\, dt.$$

Soit $x \in [0, a[$. Fixons un d tel que $x < d < a$. On a :

$$f(d) = P_n(d) + R_n(d),$$

ce qui permet d'écrire l'encadrement grossier, compte tenu du fait que les dérivées sont ≥ 0 :

$$0 \leq R_n(d) \leq f(d).$$

Par changement de variable $u = dt/x$, on obtient :

$$R_n(x) = \left(\frac{x}{d}\right)^{n+1} \int_0^d \frac{(d-u)^n}{n!} f^{n+1}(\frac{ux}{d})\, du$$

soit (puisque $f^{n+1}(\frac{ux}{d}) \leq f^{n+1}(x) \leq f^{n+1}(d)$) :

$$0 \leq R_n(x) \leq \left(\frac{x}{d}\right)^{n+1} R_n(d) \leq \left(\frac{x}{d}\right)^{n+1} f(d).$$

Le terme de droite de cette double inégalité tend bien vers 0. ■

Ceci s'applique notamment à la fonction $\tan(x)$ dont on peut ainsi montrer simplement qu'elle est développable en série entière autour de 0, le rayon de convergence de cette série étant $\pi/2$. Bien entendu, ce dernier résultat a une explication plus profonde. Il faut pour cela se placer dans le plan complexe. Soit f une fonction analytique au voisinage d'un point z_0. Le rayon de convergence de la série entière, développement de f au voisinage de z_0 est la distance de z_0 au plus proche point singulier (complexe) de f. En effet le cercle de convergence comporte toujours au moins un point singulier. Et il n'y a pas de point singulier à l'intérieur du disque de convergence. Si on veut appliquer ceci à la fonction tangente, on commence par montrer qu'elle est bien analytique tant que $\cos(z)$ ne s'annule pas, comme quotient de deux fonctions analytiques. Le point singulier le plus proche de 0 est le zéro de $\cos(z)$ le plus proche de 0 : c'est $\pi/2$.
Cette façon de voir permet de comprendre entre autres choses, pourquoi le développement de $1/(1+x^2)$ en 0 a pour rayon de convergence 1, alors même que sur $\mathbb{R}$, la fonction f est définie partout. En fait le rayon de convergence est tributaire de ce qu'il se passe dans $\mathbb{C}$ et pas seulement dans $\mathbb{R}$.

5.8.6 Retour sur la méthode de Newton

La formule de Taylor, permet de revenir sur la méthode de Newton exposée dans la section 5.4.3. En appliquant alors (pourvu que les hypothèses soient remplies) la formule de Taylor à l'ordre 2, on peut montrer que dans cette méthode l'approximation est quadratique, ce qu'on avait déjà constaté dans un certain nombre de cas particuliers.

5.9 La formule d'Euler-Maclaurin

5.9.1 Un exemple à la main

▶ **Présentation du problème**

Il s'agit d'étudier la convergence de la suite $(S_n)_{n\geq 1}$ de terme général :

$$S_n = \frac{1}{1^2} + \frac{1}{2^2} + \cdots + \frac{1}{n^2}.$$

Puis, ayant établi la convergence de cette suite vers une limite S, nous essaierons d'évaluer la rapidité de la convergence en évaluant $S - S_n$.

▶ **Convergence de la suite**

La suite $(S_n)_{n\geq 1}$ est croissante. En effet :

$$S_{n+1} - S_n = \frac{1}{(n+1)^2},$$

donc :

$$S_{n+1} \geq S_n.$$

Considérons la courbe $(\mathcal{C})$ d'équation :

$$y = \frac{1}{x^2}.$$

La somme S_n s'interprète comme la somme des aires des triangles hachurés sur la figure 5.5. En considérant par ailleurs l'aire sous la courbe, on en déduit que :

$$S_n \leq 1 + \int_1^n \frac{\mathrm{d}x}{x^2}.$$

En conséquence :

$$S_n \leq 2 - \frac{1}{n} \leq 2.$$

La suite $(S_n)_{n\geq 1}$ étant croissante et majorée converge. Elle converge vers un nombre $S \leq 2$. On peut montrer que $S = \pi^2/6$, ce que nous admettrons ici.

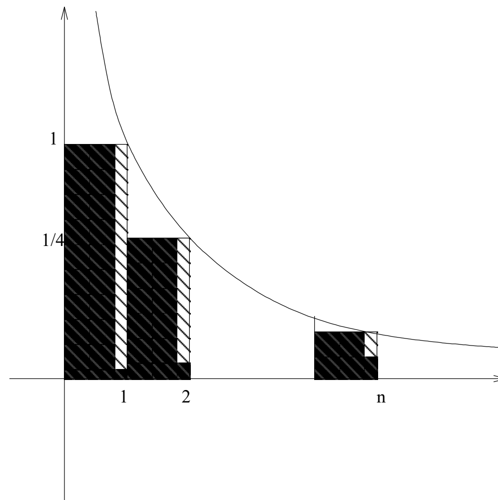


FIG. 5.5 – Majoration

▶ **Évaluation de $S - S_n$**

Posons :

$$S - S_n = R_n.$$

L'entier n étant fixé, pour tout entier m tel que $m \geq n+1$, posons :

$$R_{n,m} = \frac{1}{(n+1)^2} + \cdots + \frac{1}{m^2},$$

qu'on peut encore écrire :

$$R_{n,m} = S_m - S_n.$$

L'entier n étant fixé, la suite $(R_{n,m})_{m>n}$ converge donc vers $R_n = S - S_n$. On obtient un encadrement de $R_{n,m}$ par une majoration du style de celle suggérée par la figure 5.5 et par une minoration telle que celle suggérée par la figure 5.6.

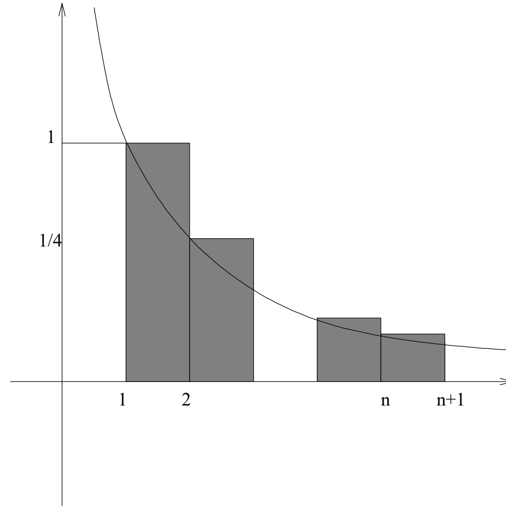


FIG. 5.6 – Minoration

$$\int_{n+1}^{m+1} \frac{\mathrm{d}x}{x^2} \leq R_{n,m} \leq \int_{n}^{m} \frac{\mathrm{d}x}{x^2},$$

c'est-à-dire que pour tout $m > n$:

$$\frac{1}{n+1} - \frac{1}{m+1} \leq R_{n,m} \leq \frac{1}{n} - \frac{1}{m},$$

et donc :

$$\frac{1}{n+1} \leq R_n \leq \frac{1}{n}.$$

Cet encadrement nous permet d'obtenir l'encadrement suivant :

$$0 \leq \frac{1}{n} - R_n \leq \frac{1}{n} - \frac{1}{n+1} \leq \frac{1}{n^2}.$$

▶ **Amélioration de l'évaluation de $S - S_n$**

On va évaluer plus finement la différence entre l'aire sous la courbe et l'aire des rectangles construits sur la figure 5.5. Pour cela notons :

$$A_k = \int_{k-1}^{k} \frac{\mathrm{d}x}{x^2} - \frac{1}{k^2}$$

et :

$$I_{n,m} = \int_{n}^{m} \frac{\mathrm{d}x}{x^2}.$$

Alors :

$$I_{n,m} - R_{n,m} = \sum_{k=n+1}^{m} A_k,$$

et en passant à la limite sur m :

$$\frac{1}{n} - R_n = \lim_{m\to\infty} \sum_{k=n+1}^{m} A_k.$$

Or :

$$A_k = \frac{1}{k-1} - \frac{1}{k} - \frac{1}{k^2} = \frac{1}{k^2(k-1)},$$

ce qui nous permet de donner l'encadrement suivant :

$$\frac{1}{k^3} \leq A_k \leq \frac{1}{(k-1)^3}.$$

En faisant cette fois-ci intervenir la courbe d'équation :

$$y = \frac{1}{x^3},$$

et en raisonnant sur des aires bien choisies comme dans le paragraphe précédent on obtient successivement :

$$\int_{k}^{k+1} \frac{\mathrm{d}x}{x^3} \leq A_k \leq \int_{k-2}^{k-1} \frac{\mathrm{d}x}{x^3},$$

$$\frac{1}{2}\left(\frac{1}{k^2} - \frac{1}{(k+1)^2}\right) \leq A_k \leq \frac{1}{2}\left(\frac{1}{(k-2)^2} - \frac{1}{(k-1)^2}\right),$$

et par sommation et simplification :

$$\frac{1}{2}\left(\frac{1}{(n+1)^2} - \frac{1}{(m+1)^2}\right) \leq \sum_{k=n+1}^{m} A_k \leq \frac{1}{2}\left(\frac{1}{(n-1)^2} - \frac{1}{(m-1)^2}\right).$$

On en conclut par passage à la limite sur m :

$$\frac{1}{2}\frac{1}{(n+1)^2} \leq \frac{1}{n} - R_n \leq \frac{1}{2}\frac{1}{(n-1)^2}.$$

On peut alors écrire :

$$\left|\frac{1}{n} - \frac{1}{2\,n^2} - R_n\right| \leq \frac{1}{2}\left(\frac{1}{(n-1)^2} - \frac{1}{(n+1)^2}\right),$$

$$\left|\frac{1}{n} - \frac{1}{2\,n^2} - R_n\right| \leq \frac{2}{(n-1)^3}.$$

On peut donc conclure que :

$$\frac{1}{n} - \frac{1}{2n^2}$$

est une bonne approximation de R_n.

Si nous analysons ce que nous avons fait :

(1) Nous avons comparé la série à une intégrale, c'est-à-dire que sur les intervalles $[k, k+1]$ nous avons d'une part l'intégrale :

$$I_{k,k+1} = \int_k^{k+1} f(t)dt,$$

(avec ici $f(x) = 1/x^2$) qui représente l'aire sous la courbe donnée par la fonction f au dessus de $[k, k+1]$, d'autre part $f(k)$ qui représente l'aire sous la marche d'escalier de hauteur $f(k)$ au dessus de $[k, k+1]$ (marche d'escalier au dessus de la courbe) ou encore éventuellement $f(k+1)$ qui représente l'aire de la marche d'escalier au dessous de la courbe.

(2) Nous avons cherché à évaluer la différence entre l'aire sous la courbe et l'aire sous une quelconque des deux marches (au dessus ou au dessous), c'est-à-dire :

$$I_{k,k+1} - f(k), \text{ ou si on veut } f(k+1) - I_{k,k+1}.$$

Nous avons fait cette évaluation de manière un peu artisanale. Nous allons voir maintenant, comment ce même problème peut être attaqué de manière plus systématique par la formule d'Euler-Maclaurin que nous allons exposer, en reprenant l'évaluation de la différence entre ces deux aires (mais pour f quelconque, suffisamment régulière).

5.9.2 Un pas vers la formule d'Euler-Maclaurin

Soit f une fonction de classe C^3 sur $[0,1]$. Cherchons à exprimer

$$\int_0^1 f(t)dt.$$

Pour cela on peut commencer par dire en remplaçant f par sa valeur en 0 qu'une valeur approchée de l'intégrale est $f(0)$. Étudions alors

$$W = \int_0^1 f(t)dt - f(0).$$

On est donc dans un cas tout à fait similaire à celui traité à la main. on voit que

$$W = \int_0^1 (f(t) - f(0))\, dt$$

ce qui par intégration par partie (on dérive $f(t) - f(0)$ et on intégre 1) donne

$$W = [P_1(t)\,(f(t) - f(0))]_0^1 - \int_0^1 f'(t)P_1(t)dt,$$

où $P_1(t)$ est une primitive de 1 (donc de la forme $t - a$) à bien choisir.

$$W = (1-a)\,(f(1) - f(0)) - \int_0^1 f'(t)P_1(t)dt.$$

L'intégrale qui reste dans le second membre peut à son tour être intégrée par partie en dérivant f' et en intégrant P_1. Soit $P_2(t)$ une primitive de $P_1(t)$. Choisissons $P_1(t)$ tel que $P_2(1) = P_2(0)$ ou encore

$$\int_0^1 P_1(t)dt = 0.$$

Ceci impose de prendre $a = 1/2$ ($P_1(t) = t - 1/2$). On a alors

$$W = 1/2\,(f(1) - f(0)) - P_2(0)\left(f'(1) - f'(0)\right) + \int_0^1 P_2(t)f^{(2)}(t)dt.$$

Intégrons de nouveau par partie l'intégrale qui subsiste au second membre. Notons $P_3(t)$ une primitive de $P_2(t)$ et choisissons $P_2(t)$ de telle sorte que $P_3(1) = P_3(0)$. Comme $P_1(t) = t - 1/2$ on a $P_2(t) = t^2/2 - t/2 + C$ et la condition imposée

$$\int_0^1 P_2(t)dt = 0$$

donne $C = 1/12$. Alors $P_3(t) = t^3/6 - t^2/4 + t/12 + C$, et si on impose aussi la condition

$$\int_0^1 P_3(t)dt = 0$$

alors $P_3(t) = t^3/6 - t^2/4 + t/12$. On obtient après une nouvelle intégration par partie :

$$W = 1/2\,(f(1) - f(0)) - 1/12\,\big(f'(1) - f'(0)\big) - \int_0^1 P_3(t) f^{(3)}(t)dt.$$

Arrêtons là le développement et écrivons en conclusion

$$\int_0^1 f(t)dt = 1/2\,(f(0) + f(1)) - 1/12\,\big(f'(1) - f'(0)\big) - \int_0^1 P_3(t) f^{(3)}(t)dt.$$

5.9.3 Polynômes de Bernoulli

Les calculs précédents mettent en évidence la suite des polynômes de Bernoulli.

Théorème 5.15 *Il existe une suite $(Q_n)_{n\geq 0}$ et une seule de polynômes telle que*

(1) $Q_0 = 1$,

(2) $Q_n' = Q_{n-1}$, *pour* $n \geq 1$,

(3) $\int_0^1 Q_n(u)du = 0$, *pour* $n \geq 1$.

Ces polynômes sont appelés ***polynômes de Bernoulli****.*

Preuve. Par récurrence sur n. Le polynôme Q_0 est bien défini et, Q_{n-1} étant construit, la condition (2) fixe Q_n à une constante près. La condition (3) fixe cette constante. ■

Compte tenu du mode de construction de ces polynômes, il est facile de voir que leurs coefficients sont rationnels.

Voici les premiers polynômes de Bernoulli :

$$Q_0 = 1 \quad Q_1 = x - \frac{1}{2} \quad Q_2 = \frac{x^2}{2} - \frac{x}{2} + \frac{1}{12}$$

$$Q_3 = \frac{x^3}{6} - \frac{x^2}{4} + \frac{x}{12}.$$

Proposition 5.1 *Pour $n \geq 2$, $Q_n(1) = Q_n(0)$.*

Preuve. En effet on doit avoir $\int_0^1 Q_{n-1}(u)du = 0$. Mais comme Q_n est une primitive de Q_{n-1} on a $\int_0^1 Q_{n-1}(u)du = Q_n(1) - Q_n(0)$, d'où le résultat. ■

Proposition 5.2 *Si* $n \geq 1$, $Q_n(x+1) - Q_n(x) = \dfrac{x^{n-1}}{(n-1)!}$.

Preuve. On constate que l'égalité est vraie pour $n = 1$. Supposons la vraie pour n, par primitivation on obtient alors l'égalité à l'ordre $n + 1$ à une constante près. Mais la proposition précédente permet de conclure à la nullité de cette constante d'intégration. ■

Cette égalité permet de calculer des sommes du type $\sum_{k=1}^{p} k^n$. Par exemple si $n = 2$ on obtient

$$Q_3(p+1) - Q_3(1) = \frac{1}{2}\sum_{k=1}^{p} k^2,$$

ce qui donne

$$\sum_{k=1}^{p} k^2 = \frac{p(p+1)(2p+1)}{6}.$$

Proposition 5.3 *Pour tout* $n \geq 0$, $Q_n(1-x) = (-1)^n Q_n(x)$.

Preuve. Le résultat est vrai pour $n = 0$ et $n = 1$. Supposons le résultat vrai pour $n \geq 2$. Alors par primitivation on obtient au rang $n+1$ la formule voulue à une constante d'intégration près

$$Q_{n+1}(1-x) = (-1)^{n+1} Q_{n+1}(x) + C.$$

Si $n + 1$ est pair en donnant à x la valeur 0 on calcule $C = 0$.
Si $n + 1$ est impair alors par primitivation

$$Q_{n+2}(1-x) = Q_{n+2}(x) + Cx + D,$$

en prenant $x = 0$ on obtient d'abord $D = 0$, puis en prenant $x = 1$ on obtient $C = 0$. ■

Proposition 5.4 *Pour* $n \geq 1$,

$$Q_{2n+1}(0) = Q_{2n+1}(1/2) = Q_{2n+1}(1) = 0.$$

Preuve. Il suffit d'appliquer la proposition précédente avec $x = 0$, ce qui donne $Q_{2n+1}(0) = Q_{2n+1}(1) = 0$, puis avec $x = 1/2$, ce qui donne $Q_{2n+1}(1/2) = 0$. ■

En étudiant par récurrence les variations des fonctions Q_n, on pourra montrer que :

Proposition 5.5 *Pour $n \geq 1$, $Q_{2n}(0) \neq 0$ et Signe$(Q_{2n}(0)) = (-1)^{n+1}$.*

On notera

$$B_k = (-1)^{k+1}\,(2k)!\,Q_{2k}(0).$$

Ainsi les B_k (**nombres de Bernoulli**) sont des rationnels positifs. Les premiers nombres de Bernoulli sont

$$B_1 = 1/6, \quad B_2 = 1/30, \quad B_3 = 1/42, \quad B_4 = 1/30.$$

5.9.4 Formule d'Euler-Maclaurin

Soit f une fonction réelle de classe C^∞ sur $[0,1]$.

$$f(1) - f(0) = \int_0^1 f'(t)dt,$$

$$f(1) - f(0) = [Q_1(t)f'(t)]_0^1 - \int_0^1 Q_1(t)f^{(2)}(t)dt,$$

$$f(1) - f(0) = 1/2(f'(1) + f'(0)) - \int_0^1 Q_1(t)f^{(2)}(t)dt,$$

et par récurrence on montre que :

Théorème 5.16 *Soit f une fonction réelle de classe C^∞ sur $[0,1]$. Alors pour tout $n \geq 1$:*

$$f(1) - f(0) = \frac{1}{2}\big(f'(1) + f'(0)\big) + \sum_{k=1}^{n} \frac{(-1)^k B_k}{(2k)!}\big(f^{(2k)}(1) - f^{(2k)}(0)\big)$$

$$- \int_0^1 Q_{2n+1}(x)f^{(2n+2)}(x)dx.$$

En particulier si on applique ce résultat à une primitive de f on obtient

$$\int_0^1 f(t)dt = \frac{1}{2}\big(f(1) + f(0)\big) + \sum_{k=1}^{n} \frac{(-1)^k}{B}{}_{k}(2k)!\big(f^{(2k-1)}(1) - f^{(2k-1)}(0)\big)$$

$$- \int_0^1 Q_{2n+1}(x)f^{(2n+1)}(x)dx.$$

On peut appliquer le théorème 5.16 sur un intervalle $[a,b]$ au lieu de $[0,1]$. Il suffit comme toujours de faire le changement de variable affine qui envoie a sur 0 et b sur 1 : $t = a + u(b-a)$. On obtient alors :

Théorème 5.17 *Soit f une fonction réelle de classe C^∞ sur $[a,b]$. Alors pour tout $n \geq 1$:*

$$f(b)-f(a) = \frac{b-a}{2}\big(f'(b)+f'(a)\big)+\sum_{k=1}^{n}\frac{(-1)^k B_k (b-a)^{2k}}{(2k)!}\big(f^{(2k)}(b)-f^{(2k)}(a)\big)$$

$$-\int_a^b Q_{2n+1}(\frac{u-a}{b-a})(b-a)^{2n-1}f^{(2n+2)}(u)du.$$

Découpons maintenant l'intervalle $[a,b]$ en q morceaux de même longueur $h=(b-a)/q$ sous la forme :

$$a = a_0 < a_1 < a_2 < \cdots < a_{q-1} < a_q = b,$$

appliquons le théorème 5.17 sur chaque morceau et sommons les résultats. Nous obtenons :

Théorème 5.18 *Soit f une fonction réelle de classe C^∞ sur $[a,b]$. Soit $a = a_0 < a_1 < \cdots < a_{q_1} < a_q = b$ le partage de $[a,b]$ en q intervalles de même longueur $h=(b-a)/q$. Alors pour tout $n \geq 1$:*

$$f(b)-f(a) = f(a_q)-f(a_0) = \frac{h}{2}(f'(b)+f'(a)) + h\sum_{s=1}^{q-1} f'(a_s)$$

$$+\sum_{k=1}^{n}\frac{(-1)^k B_k h^{2k}}{(2k)!}\big(f^{(2k)}(b)-f^{(2k)}(a)\big)$$

$$-h^{2n-1}\int_0^h Q_{2n+1}\left(\frac{v}{h}\right)\left(\sum_{s=0}^{q-1} f^{(2n+2)}(a_s+v)\right)dv.$$

Cette dernière formule est appelée la formule sommatoire d'Euler-Maclaurin car elle permet de calculer la somme :

$$\sum_{s=1}^{q-1} f'(a_s).$$

5.9.5 Application à l'évaluation de restes de séries

Considérons la série de terme général $1/n^2$. On sait que cette série converge et que :

$$S = \sum_{k=1}^{\infty}\frac{1}{k^2} = \frac{\pi^2}{6}.$$

Écrivons la somme S sous la forme :

$$S = S_p + R_p$$

où :

$$S_p = \sum_{k=1}^{p} \frac{1}{k^2} \quad \text{et} \quad R_p = \sum_{k=p+1}^{\infty} \frac{1}{k^2}.$$

Nous cherchons à évaluer R_p. Pour cela introduisons la fonction $f(x) = -1/x$ dont la dérivée est f'^2. Appliquons le théorème 5.18 à la fonction f sur l'intervalle $[p, r+1]$ (où p et r sont des entiers) avec $h = 1$ et $n = 1$. Nous obtenons successivement :

$$R_p - R_r = \frac{1}{(p+1)^2} + \frac{1}{(p+2)^2} + \cdots + \frac{1}{r^2},$$

$$R_p - R_r = \frac{1}{p} - \frac{1}{r+1} - \frac{1}{2}\left(\frac{1}{p^2} + \frac{1}{(r+1)^2}\right) + \frac{1}{6}\left(\frac{1}{p^3} - \frac{1}{(r+1)^3}\right) + T_{p,r},$$

et en faisant tendre r vers $+\infty$:

$$R_p = \frac{1}{p} - \frac{1}{2p^2} + \frac{1}{6p^3} + T_p,$$

où T_p se calcule facilement en utilisant le théorème 5.18. Un calcul simple (comparaison d'une somme de série avec une intégrale) permet de voir que :

$$T_p = O\left(\frac{1}{p^4}\right),$$

ce qui donne pour R_p le développement asymptotique :

$$R_p = \frac{1}{p} - \frac{1}{2p^2} + \frac{1}{6p^3} + O\left(\frac{1}{p^4}\right).$$

5.9.6 Remarque

Dans l'étude précédente, nous avons appliqué plusieurs méthodes conjointes :

1. tout d'abord nous sommes passé du discret au continu en comparant une série avec une intégrale ;
2. ensuite, nous avons appliqué une méthode d'intégration par partie ou plutôt une de ses conséquences évoluées : la formule d'Euler-Maclaurin.

Si nous étions resté dans le cadre strict des séries, nous aurions pu arriver (mais c'est assez pénible) à un développement asymptotique du reste de la série de terme général $1/n^2$ en utilisant la version discrète de l'intégration par partie, c'est-à-dire la transformation d'Abel.

5.10 Le théorème de Weierstrass

5.10.1 Présentation du problème

Nous montrons essentiellement les deux versions suivantes du théorème de Weierstrass.

Théorème 5.19 *Soit f une fonction continue sur un intervalle fermé borné $[a, b]$ à valeurs complexes. Il existe une suite de polynômes qui converge uniformément vers f.*

Théorème 5.20 *Soit f une fonction continue sur $\mathbb{R}$, 2π-périodique, à valeurs complexes. Il existe une suite de polynômes trigonométriques qui converge uniformément vers f.*

Nous donnons sous forme d'exercices diverses démonstrations de ces résultats. En particulier nous insistons sur l'importance de la notion de noyau positif. Nous terminons en donnant une version améliorée de ces résultats, exprimée en terme d'opérateurs positifs (Théorème de Popoviciu, Bohman et Korovkin).

5.10.2 La démonstration élémentaire d'Henri Lebesgue

▶ **Approximation de $|x|$**

Exercice 1 (Méthode 1) *Soit x un élément du segment $[-1, 1]$. Posons $u = 1 - x^2$, donc $|x| = \sqrt{1-u}$. En conclure que la fonction $f(x) = |x|$ est sur $[-1, 1]$ limite uniforme d'une suite de fonctions polynomiales.*

Exercice 2 (Méthode 2) *Définissons une suite de fonctions polynomiales à coefficients réels par*

$$\begin{cases} P_0(x) = 0 \\ P_n(x) = P_{n-1}(x) + \dfrac{1}{2}\left(x - P_{n-1}^2(x)\right) \quad \text{pour } n \geq 1 \end{cases}$$

Montrer que la suite P_n converge uniformément sur $[0, 1]$ vers $\sqrt{x}$.
En conclure que $|x|$ est sur $[-1, 1]$ limite uniforme d'une suite de fonctions polynomiales.

▶ **La méthode de Lebesgue**

Exercice 3 *Soit $C[0, 1]$ l'espace des fonctions continues sur $[0, 1]$ à valeurs dans $\mathbb{R}$, muni de la norme uniforme. Soit $\mathcal{A}$ le sous espace de $C[0, 1]$ constitué des fonctions continues affines par morceaux.*
a) *Montrer que $\overline{\mathcal{A}} = C[0, 1]$.*

b) *Montrer que les fonctions constantes et les fonctions du type* $(x - a + |x - a|)$ *forment un système générateur pour* $\mathcal{A}$.

c) *Soit* $\mathcal{P}$ *le sous espace des fonctions polynomiales de* $[0,1]$ *dans* $\mathbb{R}$. *Montrer que* $\overline{\mathcal{P}} = C[0,1]$

d) *Soit* $[a,b]$ *un intervalle fermé borné de* $\mathbb{R}$, *montrer que tout élément de* $C[a,b]$ *(espace des fonctions continues de* $[a,b]$ *dans* $\mathbb{R}$*) est limite uniforme d'une suite de fonctions polynomiales.*

e) *Démontrer un résultat analogue pour* $C_{\mathbb{C}}[a,b]$, *espace des fonctions continues de* $[a,b]$ *dans* $\mathbb{C}$.

5.10.3 Les noyaux positifs

Exercice 4 *Soit* $I_\delta(x) = \{y \mid 0 \leq y \leq 1 \text{ et } |x-y| \geq \delta\}$. *Soit* $(K_n)_n$ *une suite de fonctions de deux variables réelles à valeurs réelles telles que*

a) *Pour tout* $x \in \mathbb{R}$ *on a*

$$\int_0^1 K_n(x,y)dy = 1 \quad n = 1, 2, \cdots$$

b) *Pour tout* $x \in \mathbb{R}$ *on a*

$$\lim_{n\to\infty} \int_{I_\delta(x)} K_n(x,y)dy = 0 \quad \delta > 0.$$

c) *Les* K_n *sont positifs,*

$$K_n(x,y) \geq 0 \quad n = 1, 2, \cdots$$

Soit f *une fonction continue à valeurs complexes sur* $[0,1]$. *Définissons*

$$A_n(f)(x) = \int_0^1 K_n(x,y) f(y)dy.$$

Montrer que pour tout $x \in [0,1]$,

$$\lim_{n\to\infty} A_n(f)(x) = f(x)$$

et que si dans l'hypothèse b) la convergence est uniforme par rapport à x *alors* $A_n(f)$ *converge uniformément vers* f.

Exercice 5 *Soit* $[a,b]$ *un segment tel que* $a < 0 < b$. *On considère une suite* $(\phi_n)_n$ *de fonctions réelles intégrables sur* $[a,b]$ *satisfaisant à*

a) Pour tout entier $n > 0$ on a

$$\int_a^b \phi_n(t)dt = 1.$$

b) Pour tout $\eta > 0$ tel que $[-\eta, \eta] \subset [a, b]$

$$\lim_{n\to\infty} \left(\int_a^{-\eta} \phi_n(t)dt + \int_\eta^a \phi_n(t)dt \right) = 0.$$

c) Pour tout entier $n > 0$ on a $\phi_n \geq 0$.
Etant donnée une fonction f continue sur $\mathbb{R}$ on pose

$$\phi_n \star f(x) = \int_a^b f(x-t)\phi_n(t)dt.$$

Montrer que la suite $(\phi_n \star f)_n$ converge uniformément vers f sur tout compact.

Exercice 6 *Soit f une fonction continue périodique de période 2π. On pose*

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t)\cos(kt)dt,$$

$$b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t)\sin(kt)dt,$$

puis

$$S_n(f)(x) = \frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos(kx) + b_k \sin(kx)),$$

et

$$\sigma_n(f)(x) = \frac{1}{n} \sum_{k=0}^{n-1} S_k(f)(x).$$

Montrer que

$$S_n(f)(x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x+t) \frac{\sin\left(\left(n+\frac{1}{2}\right)t\right)}{2\sin\left(\frac{t}{2}\right)} dt$$

et que

$$\sigma_n(f)(x) = \frac{1}{n\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} f(x+2t) \left(\frac{\sin(nt)}{\sin(t)}\right)^2 dt.$$

Montrer que $\sigma_n(f)(x)$ converge uniformément vers $f(x)$ (théorème de Fejér).

Exercice 7 *Posons (noyau de Landau)*

$$Q_n(x) = \begin{cases} c_n(1-x^2)^n & si\ -1 \leq x \leq 1 \\ 0 & si\ x > 1\ ou\ x < -1. \end{cases}$$

où c_n est choisi de telle sorte que

$$\int_{-1}^{1} Q_n(x)dx = 1.$$

Montrer que si $x \in [0,1]$ alors

$$(1-x^2)^n \geq 1 - nx^2,$$

puis que $c_n \leq \sqrt{n}$.
Soit f une fonction continue sur $[0,1]$ telle que $f(0) = f(1) = 0$. Étudier

$$Q_n \star f(x) = \int_0^1 f(t)Q_n(x-t)dt$$

où $x \in [0,1]$. En conclure le théorème de Weierstrass.

Exercice 8 *Soit f une fonction continue sur $[0,1]$. Posons (polynômes de Bernstein)*

$$B_n(f)(x) = \sum_{p=0}^{n} C_n^p f\left(\frac{p}{n}\right)(1-x)^{n-p}x^p.$$

On se propose de montrer que la suite $(B_n(f))_{n\geq 1}$ converge uniformément vers f.

a) Posons $r_p(x) = C_n^p(1-x)^{n-p}x^p$, montrer que

$$\sum_{p=0}^{n} r_p(x) = 1,$$

$$\sum_{p=0}^{n} p r_p(x) = nx,$$

$$\sum_{p=0}^{n} p(p-1) r_p(x) = n(n-1)x^2.$$

(On pourra utiliser le développement de $(x+y)^n$).

b) Calculer

$$\sum_{p=0}^{n}(p-nx)^2 r_p(x).$$

c) On sait que $\forall\epsilon>0, \exists\delta>0$ *t.q.* $|x-x'|\leq\delta$ *implique* $|f(x)-f(x')|\leq\epsilon$. *On étudiera alors la somme*

$$\sum_{p=0}^{n}\left(f(x)-f\left(\frac{p}{n}\right)\right)r_p(x)$$

en la décomposant en deux suivant que $|p-nx|\leq\delta n$ *ou que* $|p-nx|>\delta n$ *et on conclura.*

5.10.4 Les opérateurs positifs

Exercice 9 *Soit* K *un espace topologique compact et* $C(K)$ *l'espace des fonctions continues sur* K *à valeurs réelles normé par*

$$\|f\| = \sup_{x\in K}|f(x)|.$$

Soit T *un opérateur linéaire de* $C(K)$ *dans lui même tel que*

$$\forall f\in C(K),\ f>0 \Longrightarrow T(f)\geq 0.$$

On dit alors que T *est un opérateur positif.*
Montrer que T *est un opérateur continu et que*

$$\|T\| = \|1_K\|,$$

où 1_K *est la fonction constante valant* 1 *sur* K.

Exercice 10 *Soit* B_n *l'opérateur de* $C[0,1]$ *dans lui même défini par*

$$B_n(f)(x) = \sum_{k=0}^{n}\left(\begin{array}{c} n\\ k\end{array}\right)x^k(1-x)^{n-k}f\left(\frac{k}{n}\right).$$

Calculer $\|B_n\|$.

Exercice 11 *Soit* f *une fonction continue périodique de période* 2π. *On reprend les notations de l'exercice* 6. *La fonction* f *peut être considérée comme élément de* $C^\star = C(T)$ *où* T *est le tore* $\mathbb{R}/(2\pi\mathbb{Z})$. *On définit donc sur* $C^\star$ *l'opérateur* S_n *par*

$$S_n(f)(x) = \frac{1}{\pi}\int_{-\pi}^{\pi}f(t)D_n(t-x)dt$$

avec

$$D_n(t) = \frac{\sin\left((2n+1)\frac{t}{2}\right)}{2\sin\left(\frac{t}{2}\right)}$$

(noyau de Dirichlet) et l'opérateur σ_n *par*

$$\sigma_n(f)(x) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)K_n(t-x)dt$$

avec

$$K_n(t) = \frac{1}{2n}\left(\frac{\sin\left(\frac{nt}{2}\right)}{\sin\left(\frac{t}{2}\right)}\right)^2$$

a) Montrer que σ_n *est un opérateur positif et que* $\|\sigma_n\| = 1$.

b) Montrer que pour tout n, S_n *n'est pas un opérateur positif. Montrer que*

$$\|S_n\| = \frac{1}{\pi}\int_{-\pi}^{\pi} |D_n(t)|dt = \frac{4}{\pi^2}\log(n) + O(1).$$

Un nombre x *étant fixé, calculer la norme de la forme linéaire* S_n^x *sur* $C^\star$ *définie par* $S_n^x(f) = S_n(f)(x)$.

Exercice 12 *(Théorème de Popoviciu, Bohman et Korovkin)*
Soit K *un espace topologique compact possédant au moins deux points et* $C(K)$ *l'espace des fonctions continues sur* K *à valeurs réelles normé par*

$$\|f\| = \sup_{x\in K} |f(x)|.$$

Soient $f_1, f_2, \cdots, f_m$ *des éléments de* $C(K)$ *vérifiant la propriété :*

$$(1)\quad \begin{cases} \textit{Il existe } m \textit{ fonctions continues réelles} \\ a_1, a_2, \cdots, a_m \textit{ définies sur } K, \textit{ telles que la fonction} \\ Q(x,y) = \sum_{i=1}^m a_i(y) f_i(x) \textit{ satisfasse à} \\ \qquad a)\ Q(x,y) \geq 0 \\ \qquad b)\ Q(x,y) = 0 \iff x = y. \end{cases}$$

Soit L_n *une suite d'opérateurs positifs de* $C(K)$ *dans lui-même vérifiant :*

pour tout $i (i = 1, \cdots, n)$ *la suite* $(L_n(f_i))_n$ *converge vers* f_i *dans* $C(K)$.

Le but du problème est de démontrer que dans ces conditions pour toute fonction $f \in C(K)$ *la suite* $(L_n(f))_n$ *converge vers* f.

Partie I. *Soit* E *le sous-espace vectoriel de* $C(K)$ *engendré par* f_1, f_2, $\cdots$, f_m.

a) Montrer que si $f \in E$ la suite $(L_n(f))_n$ converge vers f.

b) Montrer qu'il existe $g \in E$ tel que pour tout $x \in K$ on ait $g(x) > 0$.

Soit y fixé dans K. On note Q_y la fonction définie par $Q_y(x) = Q(x, y)$. On note ϕ_n la fonction définie par $\phi_n(y) = L_n(Q_y)(y)$.

c) Montrer que la suite $(\phi_n)_n$ converge uniformément vers 0.

d) Montrer qu'il existe $M_0 > 0$ tel que pour tout n on ait $\|L_n(1_K)\| \leq M_0$ où 1_K est la fonction constante valant 1.

Partie II. *Soit F un élément de $C(K \times K)$. Si $y \in K$ notons F_y la fonction définie par $F_y(x) = F(x, y)$. Supposons en outre que $F(x, x) = 0$ pour tout $x \in K$.*

a) Soit $\epsilon > 0$ montrer qu'il existe un compact $A \subset K \times K$ tel que

$$(x, y) \notin A \Longrightarrow |F(x, y)| \leq \epsilon,$$

$$m = \inf_{(x,y)\in A} Q(x, y) > 0.$$

Notons alors $M = \sup_{(x,y)\in A} |F(x, y)|$.

b) Montrer que pour tout couple $(x, y) \in K \times K$ on a

$$|F(x, y)| \leq \epsilon + \frac{M}{m} Q(x, y).$$

c) En conclure qu'il existe un entier N tel que $\forall n \geq N, \forall y \in K$ on ait

$$|L_n(F_y)(y)| \leq (M_0 + 1)\epsilon.$$

Partie III. *Soit f un élément de $C(k)$. On pose*

$$F(x, y) = f(x) - \frac{f(y)}{g(y)} g(x).$$

En utilisant les résultats précédents, montrer que la suite $(L_n(f))_n$ converge vers f.

Applications.

a) On prend $K = [0, 1]$, $m = 3$, $f_1 = 1_K$, $f_2 = x$, $f_3 = x^2$. on vérifiera avec $Q(x, y) = (y - x)^2$ que ce système satisfait bien à la condition (1). On prend $L_n = B_n$ (opérateur de Bernstein). Retrouver ainsi les résultats de l'exercice 8.

b) On prend $K = T = \mathbb{R}/(2\pi\mathbb{Z})$, $m = 3$, $f_1 = 1_K$, $f_2 = \cos(x)$, $f_3 = \sin(x)$. On vérifiera avec $Q(x, y) = 1 - \cos(x - y)$ que ce système satisfait bien à la condition (1). On prend $L_n = \sigma_n$ (opérateur de Fejér). Retrouver ainsi les résultats de l'exercice 6.

5.11 Interpolation de Lagrange

5.11.1 Introduction au problème

L'**interpolation** est un sujet très vaste lié aux questions d'**approximation** des fonctions. Très grossièrement il s'agit de trouver dans une classe fixée de fonctions (par exemple les fonctions polynomiales) un élément réalisant un certain nombre de contraintes. Souvent ces contraintes sont liées à la donnée d'une fonction f qu'on cherche à approcher par un procédé d'interpolation (par exemple la fonction cherchée doit prendre la même valeur que f en des points donnés). On se trouve alors confronté à plusieurs problèmes de natures différentes. Tout d'abord un problème algébrique, celui de trouver le ou les éléments de la classe choisie qui réalise les contraintes. Ensuite un problème d'approximation qui consiste lorsqu'on est parti d'une fonction f à mesurer la qualité de l'approximation théorique obtenue. Enfin un problème algorithmique, celui de déterminer un algorithme performant qui permette de calculer facilement et de manière aussi exacte que possible la ou les solutions.
Soient x_0, x_1, ..., x_n des nombres complexes distincts et y_0, y_1, ..., y_n des nombres complexes. Il s'agit de trouver un polynôme $P(X)$ vérifiant $P(x_k) = y_k$ pour toutes les valeurs de k comprises entre 0 et n.

5.11.2 L'aspect algébrique

▶ **Existence de solutions** Notons $\mathbb{C}[X]$ l'espace des polynômes à coefficients complexes et $\mathbb{C}_n[X]$ le sous espace des polynômes à coefficients complexes de degré inférieur ou égal à n. Considérons alors l'application linéaire T de $\mathbb{C}[X]$ dans $\mathbb{C}^{n+1}$ qui à un polynôme $P(X)$ fait correspondre $(P(x_0), P(x_1), ..., P(x_n))$.
On voit que le noyau $Ker(T)$ de l'application T est l'espace constitué des multiples du polynôme $N(X) = (X - x_0)(X - x_1)...(X - x_n)$ et qu'on peut écrire

$$\mathbb{C}[X] = \mathbb{C}_n[X] \oplus Ker(T).$$

Ceci nous montre que la restriction de T à $\mathbb{C}_n[X]$ est une bijection de $\mathbb{C}_n[X]$ sur $\mathbb{C}^{n+1}$.
Le résultat obtenu est donc le suivant :

Théorème 5.21 *Pour tout élément $(y_0, y_1, ..., y_n)$ de $\mathbb{C}^{n+1}$ il existe un polynôme $P(X)$ de degré $\leq n$ et un seul (polynôme d'interpolation de Lagrange) tel que $P(x_k) = y_k$ $(0 \leq k \leq n)$. Tout polynôme $Q(X)$ (de degré quelconque) vérifiant aussi $Q(x_k) = y_k$ $(0 \leq k \leq n)$ s'écrit sous la forme*

$$Q(X) = P(X) + K(X)N(X).$$

▶ **Ecriture sous la forme naturelle (base des monômes)** Si on cherche à trouver explicitement le polynôme de Lagrange écrit sous la forme habituelle $P(X) = a_0 + a_1X + ... + a_nX^n$ on est amené à résoudre le système de $n+1$ équations en les $n+1$ inconnues $a_0, ..., a_n$:

$$\begin{cases} a_0 + a_1x_0 + ... + a_nx_0^n & = & y_0 \\ a_0 + a_1x_1 + ... + a_nx_1^n & = & y_1 \\ \cdots\cdots\cdots\cdots & \cdots & \cdots \\ a_0 + a_1x_n + ... + a_nx_n^n & = & y_n. \end{cases}$$

Ce système est un système de **Vandermonde** dont on sait bien sûr qu'il admet une solution unique. Nous fournirons ultérieurement un algorithme pour résoudre ce système plus rapidement que ne le ferait une méthode classique comme la méthode du pivot par exemple.

▶ **Ecriture sous forme de Lagrange** Ainsi le calcul des coefficients a_i du polynôme cherché se ramène à la résolution d'un système de Vandermonde. Sans être très compliqué ceci n'est cependant pas immédiat. Or il peut se faire qu'on n'ait pas absolument besoin des coefficients a_i et qu'on puisse se satisfaire d'exprimer le polynôme d'interpolation dans une base mieux adaptée que la classique base des monômes. Dans cet ordre d'idée, introduisons les $n+1$ polynômes $L_k(X)$ (où $0 \leq k \leq n$) qui vérifient

$$\begin{cases} L_k(x_k) & = & 1 \\ L_k(x_j) & = & 0 \quad \text{si } j \neq k. \end{cases}$$

D'après l'étude qui précède le polynôme $L_k(X)$ existe et est unique. On vérifie aisément que $L_k(X)$ s'écrit explicitement sous la forme

$$L_k(X) = \frac{(X - x_0)...(X - x_{k-1})(X - x_{k+1})...(X - x_n)}{(x_k - x_0)...(x_k - x_{k-1})(x_k - x_{k+1})...(x_k - x_n)}.$$

Ces polynômes forment une base de $\mathbb{C}_n[X]$ et le polynôme d'interpolation s'exprime dans cette base

$$P(X) = \sum_{k=0}^{n} y_k L_k(X).$$

Il est intéressant de remarquer que le polynôme $L_k(X)$ s'écrit aussi

$$L_k(X) = \frac{N(X)}{(X - x_k)N'(x_k)}$$

où $N(X) = (X - x_0)(X - x_1)...(X - x_n)$.

▶ **Ecriture sous la forme de Newton** L'inconvénient des polynômes $L_k(X)$ est que chacun d'eux fait intervenir tous les points d'interpolation et donc si on rajoute un nouveau point tout calcul fait à partir des polynômes $L_k(X)$ doit être entièrement refait.
Prenons alors les polynômes $N_k(X) = (X - x_0)(X - x_1)...(X - x_{k-1})$ où $0 \leq k \leq n$ (avec $N_0 = 1$). Ces polynômes forment aussi une base de $\mathbb{C}_n[X]$ et le polynôme d'interpolation se décompose sur cette base sous la **forme de Newton**

$$P(X) = \sum_{k=0}^{n} b_k N_k(X).$$

Le problème est alors de calculer les coefficients b_k. Pour ce faire définissons les différences divisées successives des valeurs y_i par rapport aux points x_i :

$$[y_0] = y_0$$

$$[y_0, y_1, ..., y_k] = \frac{[y_1, ..., y_k] - [y_0, ..., y_{k-1}]}{x_k - x_0}.$$

Théorème 5.22 *Les coefficients de la décomposition du polynôme d'interpolation de Lagrange de degré n dans la base de Newton sont donnés par*

$$b_k = [y_0, y_1, ..., y_k]$$

où $0 \leq k \leq n$.

Preuve. La formule à démontrer est clairement vraie si on a $n = 0$. Supposons la formule vraie pour les polynômes de degré $n-1$ interpolant en n points. Soit alors $P_{n-1}(X)$ le polynôme d'interpolation de Lagrange associé aux points $x_0, x_1, ..., x_{n-1}$ et aux valeurs $y_0, y_1, ..., y_{n-1}$, $Q_{n-1}(X)$ le polynôme d'interpolation de Lagrange associé aux points $x_1, x_2, ..., x_n$ et aux valeurs $y_1, y_2, ..., y_n$ et

$$P(X) = \sum_{k=0}^{n} b_k N_k(X)$$

le polynôme d'interpolation de Lagrange associé aux points $x_0, x_1, ..., x_n$ et aux valeurs $y_0, y_1, ..., y_n$. Il est facile de voir que

$$P_{n-1}(X) = \sum_{k=0}^{n-1} b_k N_k(X)$$

si bien qu'il reste simplement en vertu de l'hypothèse de récurrence à établir la formule pour le coefficient b_n.

Pour cela définissons

$$\tilde{P}_n(X) = \frac{(X - x_0)Q_{n-1}(X) - (X - x_n)P_{n-1}(X)}{x_n - x_0}.$$

On vérifie que

$$\tilde{P}_n(x_i) = y_i$$

pour $0 \leq i \leq n$. Donc

$$\tilde{P}_n(X) = P(X).$$

En égalant les coefficients du terme de degré n dans l'expression des polynômes P_n et $\tilde{P}_n$ on obtient la relation cherchée. ■

En comparant l'expression de Newton du polynôme P_k d'interpolation en $k+1$ points avec celle de Lagrange on trouve en regardant les coefficients des termes de degré k

$$[y_0, y_1, ..., y_k] = \sum_{j=0}^{k} \frac{y_j}{N'_{k+1}(x_j)}.$$

Il est clair dans cette représentation que le rajout d'un nouveau point ne fait que rajouter un nouveau terme au polynôme, les autres termes restant identiques.

► **Interpolation de Lagrange-Sylvester** L'interpolation de Lagrange se généralise à l'interpolation de Lagrange Sylvester, où on impose comme condition au polynôme d'interpolation non seulement des valeurs en des points donnés, mais aussi les valeurs en ces mêmes points de ses quelques premières dérivées (le nombre de dérivées à valeurs imposées dépendant du point considéré).

Plus précisément : Soit s un entier et, pour tout entier i tel que $0 \leq i \leq s$, un entier $d_i \geq 1$. On pose

$$n = \sum_{0 \leq i \leq s} d_i.$$

Fixons $x_0, x_1, \cdots, x_s$ des nombres complexes distincts et pour tout $0 \leq i \leq s$ des nombres complexes $y_{i,0}, \cdots, y_{i,d_i-1}$. Alors il existe un polynôme $P(X)$ de degré $n-1$ et un seul tel que pour tout $0 \leq i \leq s$ et tout $0 \leq j \leq d_i - 1$ on ait :

$$P^{(j)}(x_i) = y_{i,j}.$$

Remarquons que si tous les d_i valent 1 on retrouve l'interpolation de Lagrange.

5.11.3 L'aspect algorithmique

▶ **Le calcul des différences divisées** L'algorithme des différences divisées est très simple. Il utilise l'écriture du polynôme d'interpolation sous la forme de Newton. Les coefficients sont alors calculés par la formule de récurrence établie précédemment

$$[y_0] = y_0$$

$$[y_0, y_1, ..., y_k] = \frac{[y_1, ..., y_k] - [y_0, ..., y_{k-1}]}{x_k - x_0}.$$

si bien que le calcul se fait conformément à la figure 5.7. Le nombre d'opérations à effectuer est en $O(n^2)$.

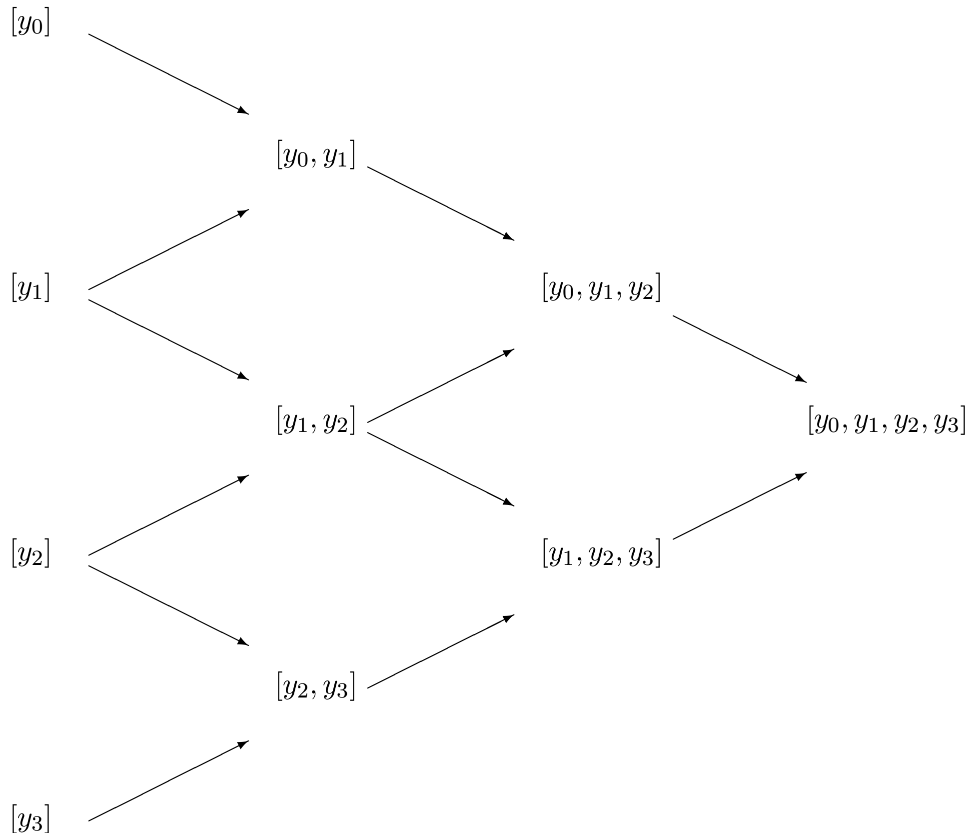


FIG. 5.7 – Le calcul des différences divisées

▶ **Résolution d'un système de Vandermonde** Nous avons vu que le calcul effectif des coefficients du polynôme d'interpolation de Lagrange dans

la base naturelle des monômes passe par la résolution d'un système de Vandermonde. Voici un algorithme qui permet de résoudre un tel système. Cet algorithme est basé en fait sur l'algorithme de Hörner pour l'évaluation de polynômes. Il est plus rapide que les algorithmes directs de résolution des systèmes linéaires généraux comme la méthode du pivot de Gauss ou la méthode de Householder qui sont en $O(n^3)$ alors que nous obtenons ici un algorithme en $O(n^2)$.

Soit N un entier ≥ 2. Etant donné $x = (x_1, x_2, ..., x_N)$ un N-uplet de réels deux à deux distincts on note B la matrice

$$B = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_N \\ x_1^2 & x_2^2 & \cdots & x_N^2 \\ \vdots & \vdots & & \vdots \\ x_1^{N-1} & x_2^{N-1} & \cdots & x_N^{N-1} \end{pmatrix}$$

On cherche à résoudre le système

$$BW = Q$$

où Q est la matrice colonne constituée des seconds membres $q_1, q_2, \cdots, q_N$ du système et où W est la matrice colonne constituée des inconnues $w_1, w_2, \cdots, w_N$ du système.

Pour tout entier j vérifiant $1 \leq j \leq N$ on pose

$$P_j(x) = \prod_{\substack{n=1 \\ n \neq j}}^{N} \frac{x - x_n}{x_j - x_n}.$$

P_j est donc un polynôme en x de degré $N-1$ qui peut s'écrire

$$P_j(x) = \sum_{k=1}^{N} A_{j,k} x^{k-1}$$

En effectuant le produit de la matrice $A = (A_{j,k})_{j,k}$ par la matrice B on constate que

$$AB = (P_j(x_k))_{j,k}$$

ce qui prouve que A est l'inverse de B. On peut alors écrire que $W = AQ$. On obtient ainsi les formules

$$w_j = \sum_{k=1}^{N} A_{j,k} q_k.$$

Nous allons dans la suite mettre en place une méthode de résolution qui calcule les coefficients des polynômes P_j, donc qui calcule l'inverse de la matrice B. Pour calculer les coefficients de P_j on sera amené à calculer les coefficients de

$$N_j(x) = \prod_{\substack{n=1 \\ n \neq j}}^{N} (x - x_n)$$

et aussi le dénominateur intervenant dans la formule qui définit P_j, c'est-à-dire le nombre $N_j(x_j)$.

Posons $P(x) = (x - x_1)(x - x_2)...(x - x_N)$. $P(x)$ est donc un polynôme de degré N qui s'écrit sous la forme :

$$P(x) = x^N + c_N x^{N-1} + ... + c_2 x + c_1.$$

Montrons tout d'abord comment si on connaît les coefficients c_j on peut calculer les coefficients du polynôme

$$N_j(x) = \prod_{\substack{n=1 \\ n \neq j}}^{N} (x - x_n).$$

Pour cela posons

$$N_j(x) = b_N x^{N-1} + ... + b_2 x + b_1.$$

On vérifie immédiatement sur l'expression de $N_j(x)$ que le coefficient du terme de plus haut degré est 1. En remarquant que $P(x) = N_j(x)(x - x_j)$ on établit la formule

$$b_{k-1} = c_k + x_j b_k.$$

Si bien que

$$\begin{cases} b_N = 1 \\ b_{k-1} = c_k + x_j b_k. \end{cases}$$

Connaissant les coefficients de N_j il est alors facile de calculer le dénominateur $N_j(x_j)$ intervenant dans la définition de P_j.
En effet posons $t_N = b_N = 1$ et définissons pour tout $k \leq N$

$$t_{k-1} = x_j t_k + b_{k-1}.$$

On constate alors que $t_1 = N_j(x_j)$, le calcul proposé pour $N_j(x_j)$ n'étant rien d'autre que l'algorithme de Horner.

Il reste maintenant à calculer les coefficients c_j de P.

Pour tout entier k vérifiant $1 \leq k \leq N$ on définit

$$Q_k(x) = (x - x_1)(x - x_2)...(x - x_k)$$

et on écrit Q_k sous la forme

$$Q_k(x) = x^k + \alpha_{k,k}x^{k-1} + \alpha_{k,k-1}x^{k-2} + ... + \alpha_{k,1}.$$

Il est facile de voir sur l'expression de $Q_1(x) = x - x_1$ que $\alpha_{1,1} = -x_1$.
De la formule $Q_k(x) = Q_{k-1}(x)(x - x_k)$ découlent pour $k = 2, 3, ..., N$ les formules

$$\alpha_{k,k} = \alpha_{k-1,k-1} - x_k$$

$$\alpha_{k,j} = \alpha_{k-1,j-1} - x_k\alpha_{k-1,j} \quad j = k - 1, ..., 2$$

ce qui achève l'algorithme.

On peut voir que le nombre d'opérations à faire dans cet algorithme est en $O(N^2)$, la partie la plus coûteuse étant le calcul des coefficients c_k.

5.11.4 Partie approximation

▶ Un exemple

On considère la fonction f définie sur l'intervalle $[0, 6]$ par l'équation :

$$y = f(x) = 6\left(\sin\left(\frac{\pi(x-3)}{6}\right) + 1\right). \tag{5.7}$$

Remarque 5.1 *Cette fonction peut représenter l'élévation du niveau de la mer pour une marée en fonction du temps. La variable x est alors l'heure marée (1/6 du temps de passage de la basse mer à la haute mer) tandis que y représente le nombre de $1/12^e$ (1/12 de la différence de niveau exprimé en mètre entre la haute mer et la basse mer). Nous allons chercher un polynôme qui approche cette fonction sinus sur cet intervalle. Le point d'inflexion nous donne à penser qu'il faut tenter une cubique pour être « dans la forme ».*

Nous allons donc chercher le polynôme d'interpolation P de degré ≤ 3 tel que :

- $P(0) = f(0) = 0$;
- $P(3) = f(3) = 6$;
- $P(6) = f(6) = 12$;
- $P'(0) = f'(0) = 0$.

Les conditions $P(0) = 0$ et $P'(0) = 0$ imposent que : $P(x) = x^2(ax + b)$. Les deux autres conditions donnent respectivement $9a + 3b = 2$ et $18a + 3b = 1$, d'où :

$$\begin{cases} a = -1/9 \\ b = 1. \end{cases}$$

Le polynôme cherché est donc : $P(x) = -\frac{1}{9}x^3 + x^2$. Nous avons représenté le graphe de P sur la figure 5.8.

▶ Qualité des approximations

Afin d'évaluer la précision de l'approximation $P(x)$ de la fonction $f(x)$, nous allons majorer la quantité

$$\sup_{x \in [0,6]} |f(x) - P(x)|,$$

c'est-à-dire la norme uniforme $||f - P||_\infty$ sur l'intervalle $[0, 6]$.

▶ Un outil fondamental : le théorème de division des fonctions différentiables

Théorème 5.23 *Soit f une fonction réelle définie sur $\mathbb{R}$ de classe C^{p+1}, où p est un entier naturel. On suppose que f s'annule en un point a de $\mathbb{R}$. Alors il existe une unique fonction continue $g(x)$ telle que :*

$$f(x) = (x - a)g(x).$$

Cette fonction g est de classe C^p et pour tout $0 \leq q \leq p$:

$$|g^{(q)}(x)| \leq \frac{1}{q+1} \sup_{t \in [a,x]} |f^{(q+1)}(t)|. \tag{5.8}$$

Preuve. La fonction g est nécessairement définie par :

$$g(x) = \begin{cases} \dfrac{f(x)}{x-a} & \text{si } x \neq a, \\ f'(a) & \text{si } x = a. \end{cases} \tag{5.9}$$

On constate en distinguant le cas où $x = a$ de celui où $x \neq a$ que :

$$g(x) = \int_0^1 f'(a + (x - a)u)du.$$

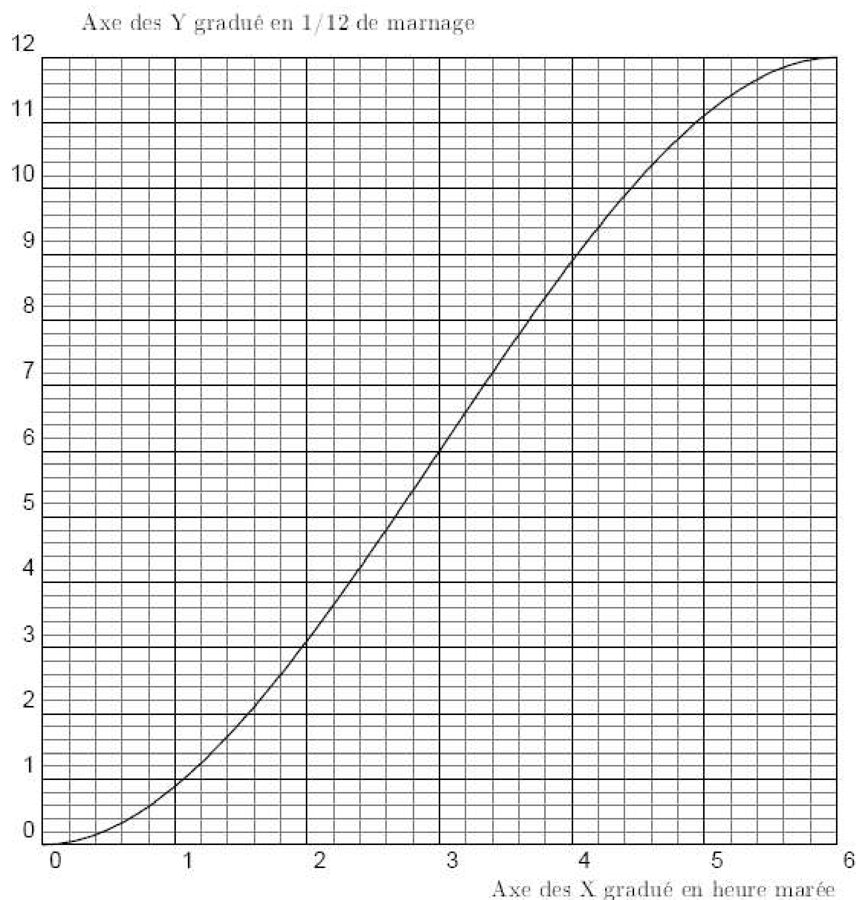


FIG. 5.8 – Approximation cubique

Sous cette dernière forme, d'après le théorème de dérivation sous le signe intégrale, on voit que la fonction $x \mapsto g(x)$ est de classe C^p et que pour tout $0 \leq q \leq p$,

$$g^{(q)}(x) = \int_0^1 u^q f^{(q+1)}(a + (x-a)u)du,$$

ce qui nous donne la formule (5.8). ■

▶ **Majoration de l'erreur dans une interpolation**

Soit f une fonction de classe C^{n+1}, P le polynôme d'interpolation de Lagrange qui prend les mêmes valeurs que f aux points $x_0, x_1, ..., x_n$ et I un intervalle

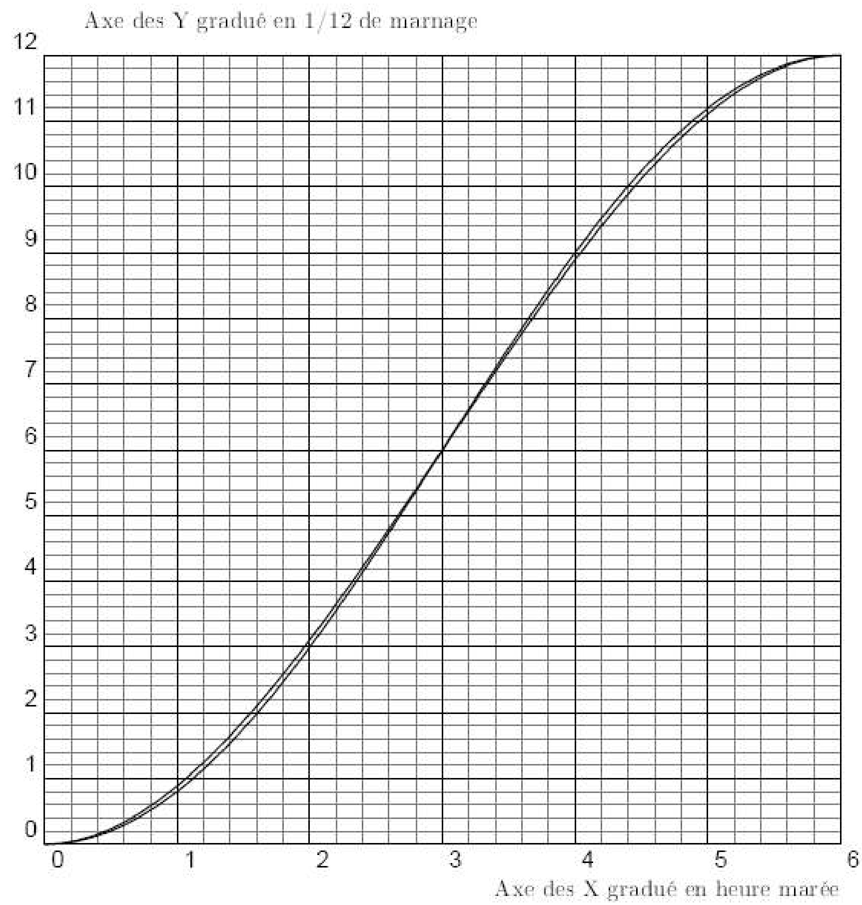


FIG. 5.9 – Comparaison entre le sinus et la cubique

compact contenant $x, x_0, x_1, ..., x_n$.
Appliquons alors le théorème 5.23 à $f(x) - P(x)$. On obtient

$$f(x) - P(x) = (x - x_0)g_0(x)$$

avec

$$|g_0^{(n)}(x)| \leq \frac{1}{n+1} \sup_{t \in I} |f^{(n+1)}(t)|$$

(ne pas oublier que $P^{(n+1)}(x) = 0$), puis

$$g_0(x) = (x - x_1)g_1(x)$$

avec

$$|g_1^{(n-1)}(x)| \leq \frac{1}{n} \sup_{t \in I} |g_0^{(n)}(t)|,$$

et ainsi de suite. Si bien que :

$$|f(x) - P(x)| \leq \frac{1}{(n+1)!} |(x - x_0)(x - x_1)...(x - x_n)| \sup_{t \in I} |f^{(n+1)}(t)|. \quad (5.10)$$

Remarque 5.2 *Cette démonstration permet d'établir de la même façon une majoration dans le cas d'une interpolation de Lagrange-Sylvester.*

▶ **Les majorations**

Si on ne regarde pas de plus près, on pourrait croire que nous avons fait une interpolation de Lagrange-Sylvester par un polynôme de degré 3 sous les conditions d'interpolation :

$$P(0) = f(0) \quad P(3) = f(3) \quad P(6) = f(6) \quad P'(0) = f'(0).$$

Mais en fait on constate qu'avec ce même polynôme de degré 3 on a aussi $P'(6) = f'(6)$. Donc on a effectué en fait une interpolation de degré 4, dont le coefficient du terme de plus haut degré est nul. On peut donc appliquer la majoration (5.10) et la remarque 5.2 à cet ordre et on obtient :

$$|f(x) - P(x)| \leq \frac{1}{5!} |x^2(x-3)(x-6)^2| \sup_{t \in [0,6]} |f^{(5)}(t)|.$$

Ceci nous donne :

$$|f(x) - P(x)| \leq \frac{1}{120} \times 6 \times \left(\frac{\pi}{6}\right)^5 |x^2(x-3)(x-6)^2|,$$

$$|f(x) - P(x)| \leq \frac{1}{120} \times 6 \times \frac{1}{25} |x^2(x-3)(x-6)^2|.$$

Pour étudier la borne supérieure de la fonction $|x^2(x-3)(x-6)^2|$ qui est symétrique par rapport à $x = 3$, il suffit de chercher son maximum sur $[0, 3]$. Sur cet intervalle cette fonction est $x^2(3-x)(x-6)^2$ et sa dérivée s'annule pour 0 et $\frac{15-3\sqrt{5}}{5}$. Le maximun est atteint en $\frac{15-3\sqrt{5}}{5}$ et est majoré par 70. En conséquence :

$$|f(x) - P(x)| \leq \frac{1}{120} \times 6 \times \frac{1}{25} \times 70,$$

$$|f(x) - P(x)| \leq 0.14$$

Remarque 5.3 *Un calcul à la machine montre qu'il semble que le maximum soit proche de* 0.12 *(pour* $x \approx 1.7$*).*

Chapitre 6

Accélération de convergence

Accélération de la convergence des suites réelles

(Jean-Etienne Rombaldi [1])

Résumé : Jean-Etienne Rombaldi nous propose un exposé clair et bien structuré sur la rapidité de convergence des suites et sur quelques procédés d'acélération de convergence. Agrémenté de 19 exercices proposés avec leurs solutions, ce travail représente un moyen de s'entraîner sur ce thème pour les écrits des concours et d'aborder une leçon d'oral du CAPES externe encore présente à la session 2008 des épreuves (et dont le libellé exact est : "Exemples d'étude de la rapidité de la convergence d'une suite réelle (u_n) vers une limite ℓ : cas où $|u_n - \ell|$ est dominé par n^{-a}, par q^n... L'exposé pourra être illustré par un ou des exemples faisant appel à l'utilisation d'une calculatrice.").

Dans toute la suite, on désigne par $(u_n)_{n\in\mathbb{N}}$ une suite de réels qui converge vers un réel ℓ. On suppose de plus que $u_n \neq \ell$ pour tout $n \in \mathbb{N}$. Les résultats classiques sur les séries numériques et les intégrales généralisées sont supposés acquis.

6.1 Vitesse de convergence

Définition 6.1 *Si la suite* $\left(\left|\dfrac{u_{n+1}-\ell}{u_n-\ell}\right|\right)_{n\in\mathbb{N}}$ *est convergente de limite* λ, *on dit que la convergence de la suite* $(u_n)_{n\in\mathbb{N}}$ *vers* ℓ *est :*

- lente, si $\lambda = 1$ *;*

[1] Université de Grenoble, jean-etienne.rombaldi@ujf-grenoble.fr.

- *géométrique de rapport* λ, *si* $\lambda \in \left]0,1\right[$;
- *rapide si* $\lambda = 0$.

Dans le cas où $\lambda \in \left]0,1\right[$, pour $\varepsilon > 0$ tel que $0 < \lambda + \varepsilon < 1$, on peut trouver un entier naturel n_ε tel que :

$$\forall n \geq n_\varepsilon,\ 0 < \left|\frac{u_{n+1} - \ell}{u_n - \ell}\right| < \lambda + \varepsilon$$

ce qui entraîne, pour $n \geq n_\varepsilon$:

$$|u_n - \ell| < (\lambda + \varepsilon)^{n - n_\varepsilon} |u_{n_\varepsilon} - \ell| = \frac{|u_{n_\varepsilon} - \ell|}{(\lambda + \varepsilon)^{n_\varepsilon}} (\lambda + \varepsilon)^n$$

ce qui signifie que la suite $(|u_n - \ell|)_{n \geq n_\varepsilon}$ est dominée par la suite géométrique $((\lambda + \varepsilon)^n)_{n \geq n_\varepsilon}$.

Définition 6.2 *On dit que le réel* λ, *quand il existe, est le coefficient de convergence de la suite* $(u_n)_{n \in \mathbb{N}}$.

Remarque 6.1 *Si la suite* $\left(\left|\frac{u_{n+1} - \ell}{u_n - \ell}\right|\right)_{n \in \mathbb{N}}$ *converge, sa limite* λ *est alors nécessairement dans* $[0,1]$. *En effet, dans le cas contraire, on a*

$$\lim_{n \to +\infty} \left|\frac{u_{n+1} - \ell}{u_n - \ell}\right| = \lambda > 1,$$

ce qui entraîne $\lim_{n \to +\infty} |u_n - \ell| = +\infty$ *et la suite* $(u_n - \ell)_{n \in \mathbb{N}}$ *est divergente (cette suite est minorée par une suite géométrique divergente), ce qui est en contradiction avec la convergence de* $(u_n)_{n \in \mathbb{N}}$ *vers* ℓ.

Remarque 6.2 *Dans la pratique, on ne connaît pas toujours la limite de la suite* $(u_n)_{n \in \mathbb{N}}$, *mais dans certains cas, on peut calculer le coefficient de convergence* λ *sans connaître explicitement cette limite* ℓ.

Lemme 6.1 *Si* $\lim_{n \to +\infty} \frac{u_{n+1} - \ell}{u_n - \ell} = \lambda \in \left]-1,1\right[\setminus \{0\}$ *(la convergence de la suite* $(u_n)_{n \in \mathbb{N}}$ *vers* ℓ *est donc géométrique de rapport* $|\lambda|$*) et s'il existe un entier* $n_0 \geq 1$ *tel que* $u_n \neq u_{n-1}$ *pour tout* $n \geq n_0$, *alors la suite* $\left(\frac{u_{n+1} - u_n}{u_n - u_{n-1}}\right)_{n \geq n_0}$ *converge vers* λ.

Preuve. Pour tout $n \geq n_0$, on a :

$$\begin{aligned}\frac{u_{n+1}-u_n}{u_n-u_{n-1}} &= \frac{u_{n+1}-\ell-(u_n-\ell)}{u_n-\ell-(u_{n-1}-\ell)} \\ &= \frac{u_n-\ell}{u_{n-1}-\ell}\times\frac{\dfrac{u_{n+1}-\ell}{u_n-\ell}-1}{\dfrac{u_n-\ell}{u_{n-1}-\ell}-1} \underset{n\to+\infty}{\to} \lambda\times\frac{\lambda-1}{\lambda-1}=\lambda\end{aligned}$$

(de $u_n \neq u_{n-1}$, on déduit que $\dfrac{u_n-\ell}{u_{n-1}-\ell} \neq 1$). ■

Remarque 6.3 *Le fait que* $\lim_{n\to+\infty}\left|\dfrac{u_{n+1}-\ell}{u_n-\ell}\right| = \lambda$ *n'entraîne pas nécessairement la convergence de la suite* $\left(\left|\dfrac{u_{n+1}-u_n}{u_n-u_{n-1}}\right|\right)_{n\geq n_0}$. *Par exemple pour la suite* $(u_n)_{n\in\mathbb{N}}$ *définie par* $u_{2p} = (-1)^p\lambda^{2p}$ *et* $u_{2p+1} = (-1)^p\lambda^{2p+1}$ *avec* $0<\lambda<1$, *on a* $|u_n| = \lambda^n \underset{n\to+\infty}{\to} \ell = 0$, $\left|\dfrac{u_{n+1}-\ell}{u_n-\ell}\right| = \lambda$, $u_n \neq u_{n-1}$ *pour tout* $n\geq 1$ *et :*

$$\frac{u_{2p+1}-u_{2p}}{u_{2p}-u_{2p-1}} = \frac{(-1)^p\lambda^{2p+1}-(-1)^p\lambda^{2p}}{(-1)^p\lambda^{2p}-(-1)^{p-1}\lambda^{2p-1}} = \frac{\lambda(\lambda-1)}{\lambda+1}$$

$$\frac{u_{2p+2}-u_{2p+1}}{u_{2p+1}-u_{2p}} = \frac{(-1)^{p+1}\lambda^{2p+2}-(-1)^p\lambda^{2p+1}}{(-1)^p\lambda^{2p+1}-(-1)^p\lambda^{2p}} = \frac{\lambda(\lambda+1)}{1-\lambda}$$

et la suite $\left(\left|\dfrac{u_{n+1}-u_n}{u_n-u_{n-1}}\right|\right)_{n\geq 1}$ *est divergente.*

Exercice 6.1 *Étudier la vitesse de convergence des suites* $(u_n)_{n\geq 2}$ *définies par* $u_n = 1/n^b$ *où* $b>0$, $u_n = 1/\ln(n)$, $u_n = a^n$ *où* $0<|a|<1$, $u_n = 1/n!$ *et* $u_n = n!/n^n$.

Solution 6.1 *Chacune de ces suites converge vers* 0 *et :*

- *pour* $u_n = 1/n^b$, *on a :*

$$\left|\frac{u_{n+1}}{u_n}\right| = \left(\frac{n}{n+1}\right)^b \underset{n\to+\infty}{\to} 1,$$

donc la convergence est lente ;

- *pour* $u_n = 1/\ln(n)$, *on a :*

$$\left|\frac{u_{n+1}}{u_n}\right| = \frac{\ln(n)}{\ln(n+1)} = \ln\left(\frac{n}{n+1}\right)\frac{1}{\ln(n+1)}+1 \underset{n\to+\infty}{\to} 1,$$

donc la convergence est lente ;

- *pour* $u_n = a^n$*, on a :*

$$\left|\frac{u_{n+1}}{u_n}\right| = |a| \underset{n\to+\infty}{\to} \lambda = |a|,$$

donc la convergence est géométrique de rapport $|a|$ *;*

- *pour* $u_n = 1/n!$*, on a :*

$$\left|\frac{u_{n+1}}{u_n}\right| = \frac{1}{n+1} \underset{n\to+\infty}{\to} 0,$$

donc la convergence est rapide ;

- *pour* $u_n = n!/n^n$*, on a :*

$$\left|\frac{u_{n+1}}{u_n}\right| = \left(\frac{n}{n+1}\right)^n \underset{n\to+\infty}{\to} \frac{1}{e} < 1,$$

donc la convergence est géométrique de rapport $\frac{1}{e}$*.*

D'un point de vue pratique, on peut utiliser les critères suivants où l'on compare la suite $(|u_n - \ell|)_{n\in\mathbb{N}}$ aux suites $\left(\frac{1}{n^b}\right)_{n\geq 1}$, $\left(\frac{1}{\ln(n)}\right)_{n\geq 2}$ ou $(\lambda^n)_{n\in\mathbb{N}}$:

- si $|u_n - \ell| \underset{+\infty}{\backsim} \frac{C}{n^b}$ où C et b sont des réels strictement positif, alors la convergence est lente ;

- si $|u_n - \ell| \underset{+\infty}{\backsim} \frac{C}{\ln(n)}$ où C est un réel strictement positif, alors la convergence est lente ;

- si $|u_n - \ell| \underset{+\infty}{\backsim} C\lambda^n$, où $C > 0$ et $\lambda \in \left]0,1\right[$, alors la convergence est géométrique de rapport λ.

Exercice 6.2 *En considérant la suite définie par :*

$$u_n = \begin{cases} \frac{1}{n} & si\ n = 2p \\ \frac{2}{n} & si\ n = 2p+1, \end{cases}$$

montrer qu'une suite convergente n'a pas nécessairement de vitesse de convergence.

Solution 6.2 *Avec $|u_n| \leq \frac{2}{n}$ pour tout $n \geq 1$, on voit que cette suite converge vers 0 et avec :*

$$\left|\frac{u_{n+1}}{u_n}\right| = \begin{cases} \dfrac{2n}{n+1} \text{ si } n = 2p \\ \dfrac{n}{2(n+1)} \text{ si } n = 2p+1, \end{cases}$$

on voit que la suite $\left(|\frac{u_{n+1}}{u_n}|\right)_{n\in\mathbb{N}}$ est divergente.

Remarque 6.4 *L'exemple précédent montre également qu'une majoration du type $|u_n - \ell| \leq \dfrac{C}{n^b}$ ne permet pas nécessairement d'avoir des informations sur la vitesse de convergence de la suite u.*
De même en considérant la suite définie par :

$$u_n = \begin{cases} \lambda^n \text{ si } n = 2p \\ 2\lambda^n \text{ si } n = 2p+1 \end{cases}$$

où $0 < \lambda < 1$, on a $|u_n| \leq 2\lambda^n \underset{n\to+\infty}{\to} 0$ et :

$$\left|\frac{u_{n+1}}{u_n}\right| = \begin{cases} 2\lambda \text{ si } n = 2p \\ \dfrac{\lambda}{2} \text{ si } n = 2p+1 \end{cases}$$

n'a pas de limite.

Exercice 6.3 *Soit $(u_n)_{n\in\mathbb{N}^*}$ la suite définie par :*

$$\forall n \geq 1,\ u_n = \left(1 + \frac{1}{n}\right)^n.$$

Montrer que la convergence de cette suite (vers le nombre e) est lente et que la convergence de la suite $(v_n)_{n\geq 0} = (u_{2^n})_{n\geq 0}$ est géométrique.

Solution 6.3 *Un développement limité à l'ordre 2 nous donne :*

$$\forall n \geq 1,\ u_n = e\left(1 - \frac{1}{2n} + o\left(\frac{1}{n}\right)\right),$$

ce qui entraîne $|e - u_n| \underset{+\infty}{\backsim} \dfrac{e}{2n}$ et la convergence de cette suite est lente.
Pour la suite $(v_n)_{n\geq 0} = (u_{2^n})_{n\geq 0}$, on a $|e - v_n| \underset{+\infty}{\backsim} \dfrac{e}{2^{n+1}}$ et la convergence est géométrique de rapport $\dfrac{1}{2}$.

De manière plus générale, dès qu'on a un développement asymptotique de la forme :

$$u_n = \ell + \beta\lambda^n + o(\lambda^n)$$

avec β non nul et $|\lambda|$ dans $]0,1[$, la convergence de la suite $(u_n)_{n\in\mathbb{N}}$ vers ℓ est géométrique de rapport $|\lambda|$.

En considérant les suites $(n\lambda^n)_{n\in\mathbb{N}}$ ou $\left(\frac{\lambda^n}{n}\right)_{n\in\mathbb{N}^*}$, on constate que la réciproque est fausse.

Il ne faut pas croire au vu de l'exemple précédent que l'on peut toujours accélérer la convergence d'une suite (on précisera cette notion au paragraphe suivant) par extraction.

Par exemple les suites $u = \left(\frac{1}{\ln(n)}\right)_{n\geq 2}$, $v = \left(\frac{1}{\ln(2^n)}\right)_{n\geq 1} = \left(\frac{1}{n\ln(2)}\right)_{n\geq 1}$ et $w = \left(\frac{1}{\ln(n^2)}\right)_{n\geq 2} = \left(\frac{1}{2\ln(n)}\right)_{n\geq 2}$ convergent toutes lentement.

Par contre un développement asymptotique de la forme :

$$u_n = \ell + \frac{\beta}{n^b} + o\left(\frac{1}{n^b}\right)$$

avec β non nul et $b > 0$ qui assure une convergence lente donne :

$$u_{2^n} = \ell + \frac{\beta}{2^{nb}} + o\left(\frac{1}{2^{nb}}\right)$$

qui assure une convergence géométrique de rapport $\frac{1}{2^b}$ de la suite $(u_{2^n})_{n\in\mathbb{N}}$.

Exercice 6.4 *Montrer que la convergence de la suite* $(u_n)_{n\in\mathbb{N}}$ *définie par :*

$$\forall n \geq 0, \quad u_n = \sum_{k=0}^{n} \frac{1}{k!},$$

est rapide.

Solution 6.4 *On sait que la suite* $(u_n)_{n\in\mathbb{N}}$ *converge vers* e. *La formule de Taylor-Lagrange nous dit que pour tout* $n \geq 1$, *il existe un réel* $c_n \in]0,1[$ *tel que :*

$$e - u_n = \frac{e^{c_n}}{n!}$$

ce qui donne :

$$0 < \frac{e - u_{n+1}}{e - u_n} = \frac{1}{n+1}\frac{e^{c_{n+1}}}{e^{c_n}} < \frac{1}{n+1}e \underset{n\to+\infty}{\to} 0$$

et la convergence est rapide.

Exercice 6.5 *Montrer que, pour tout réel $\alpha > 1$, la convergence de la suite $(u_n)_{n\geq 1}$ définie par :*

$$\forall n \geq 1, \quad u_n = \sum_{k=1}^{n} \frac{1}{k^\alpha}$$

est lente.

Solution 6.5 *On sait que la suite $(u_n)_{n\geq 1}$ converge vers un réel ℓ (série de Riemann). Pour $n \geq 1$, on a :*

$$\ell - u_n = \ell - \sum_{k=1}^{n} \frac{1}{k^\alpha} = \sum_{k=n+1}^{+\infty} \frac{1}{k^\alpha}.$$

Avec les encadrements :

$$\frac{1}{(k+1)^\alpha} \leq \int_k^{k+1} \frac{1}{t^\alpha} dt \leq \frac{1}{k^\alpha}$$

on déduit que :

$$\forall n \geq 2,\ \ell - u_n \leq \int_n^{+\infty} \frac{1}{t^\alpha} dt = \frac{1}{\alpha - 1} \frac{1}{n^{\alpha-1}} \leq \ell - u_{n-1},$$

ou encore :

$$\forall n \geq 1,\ \frac{1}{(n+1)^{\alpha-1}} \leq (\alpha - 1)(\ell - u_n) \leq \frac{1}{n^{\alpha-1}},$$

ce qui donne :

$$\ell - u_n \underset{+\infty}{\backsim} \frac{1}{\alpha - 1} \frac{1}{n^{\alpha-1}}$$

et la convergence est lente.

Exercice 6.6 *Montrer que convergence de la suite $u = (u_n)_{n\in\mathbb{N}}$ définie par $u_n = \sum\limits_{k=1}^{n} \frac{1}{k} - \ln(n+1)$ pour tout $n \geq 1$ est lente.*

Solution 6.6 *On sait que la suite $(u_n)_{n\geq 1}$ converge vers γ (constante gamma d'Euler). Pour tout $n \geq 1$, on a :*

$$u_n = \sum_{k=1}^{n} \left(\frac{1}{k} - \int_k^{k+1} \frac{dt}{t} \right) = \sum_{k=1}^{n} \int_k^{k+1} \left(\frac{1}{k} - \frac{1}{t} \right) dt = \sum_{k=1}^{n} \int_k^{k+1} \frac{t-k}{kt} dt,$$

ce qui fait apparaître γ comme somme d'une série, soit :

$$\gamma = \sum_{n=1}^{+\infty} \int_n^{n+1} \frac{t-n}{nt} dt$$

et :

$$\gamma - u_n = \sum_{k=n+1}^{+\infty} \int_k^{k+1} \frac{t-k}{kt} dt.$$

En utilisant les inégalités :

$$\int_k^{k+1} \frac{t-k}{kt} dt \leq \frac{1}{k^2} \int_k^{k+1} (t-k)\, dt = \frac{1}{2k^2} \leq \frac{1}{2} \left(\frac{1}{k-1} - \frac{1}{k} \right)$$

et :

$$\int_k^{k+1} \frac{t-k}{kt} dt \geq \frac{1}{k(k+1)} \int_k^{k+1} (t-k)\, dt = \frac{1}{2} \frac{1}{k(k+1)} = \frac{1}{2} \left(\frac{1}{k} - \frac{1}{k+1} \right)$$

on en déduit que :

$$\frac{1}{2(n+1)} < \gamma - u_n < \frac{1}{2n}$$

et $\gamma - u_n \underset{+\infty}{\sim} \dfrac{1}{2n}$. *La convergence est donc lente.*

Exercice 6.7 *Soient $I = [a, b]$ un intervalle réel fermé non réduit à un point et $f : I \to I$ de classe $\mathcal{C}^1$ telle que $0 < |f'(x)| < 1$ pour tout $x \in I$.*

1. *Montrer que f admet un unique point fixe $\ell \in I$.*
2. *Pour u_0 donné dans $I \setminus \{\ell\}$, on définit la suite $u = (u_n)_{n \in \mathbb{N}}$ en posant $u_{n+1} = f(u_n)$ pour tout $n \in \mathbb{N}$. Montrer que cette suite converge vers ℓ et que la convergence est géométrique de rapport $|f'(\ell)|$. Dans cette situation, on dit que ℓ est un point fixe attractif de f.*

Solution 6.7 *1. Avec la continuité de f et $f(I) \subset I$, on déduit que f admet au moins un point fixe. En effet, la fonction g définie sur I par $g(x) = f(x) - x$ étant continue telle que $g(a) = f(a) - a \geq 0$ et $g(b) = f(b) - b \leq 0$ (puisque $f(a)$ et $f(b)$ sont dans $I = [a, b]$), le théorème des valeurs intermédiaires nous dit qu'elle s'annule sur I. De plus l'hypothèse $-1 < f' < 1$ entraîne $g' = f' - 1 < 0$ sur I, c'est-à-dire que g est strictement décroissante sur I, donc injective, et la solution ℓ de $g(x) = 0$ sur I est unique. Ce réel ℓ est l'unique point fixe de f sur I.*

2. Si $u_0 \neq \ell$ alors $u_n \neq \ell$ pour tout $n \in \mathbb{N}$. En effet, le résultat est vrai

pour $n = 0$ et en le supposant vrai pour $n \geq 0$, le théorème des accroissements finis nous permet d'écrire $u_{n+1} - \ell = f(u_n) - f(\ell) = (u_n - \ell) f'(c_n)$ avec c_n strictement compris entre u_n et ℓ et $f'(c_n) \neq 0$, ce qui entraîne $u_{n+1} \neq \ell$.
Le développement limité qui précède nous permet également de déduire que :

$$\frac{u_{n+1} - \ell}{u_n - \ell} = f'(c_n) \underset{n \to +\infty}{\to} f'(\ell) \in \left]-1, 1\right[\setminus \{0\},$$

ce qui implique que la suite u converge vers ℓ et que la convergence est géométrique de rapport $|f'(\ell)|$.

Définition 6.3 *Soit $r \geq 2$ un entier naturel. On dit que la convergence de la suite $(u_n)_{n \in \mathbb{N}}$ vers ℓ est d'ordre r s'il existe une constante $\lambda \neq 0$ telle que*

$$\lim_{n \to +\infty} \frac{|u_{n+1} - \ell|}{|u_n - \ell|^r} = \lambda.$$

Définition 6.4 *Une convergence lente ou géométrique est dite d'ordre* 1. *On dit aussi que la convergence est super-linéaire si elle est d'ordre $r \geq 2$.*

▶ Si la convergence de $(u_n)_{n \in \mathbb{N}}$ vers ℓ est d'ordre $r \geq 1$, alors cet entier r est uniquement déterminé. En effet si

$$\lim_{n \to +\infty} \frac{|u_{n+1} - \ell|}{|u_n - \ell|^r} = \lambda \quad \text{et} \quad \lim_{n \to +\infty} \frac{|u_{n+1} - \ell|}{|u_n - \ell|^s} = \mu$$

avec $s > r \geq 1$, alors

$$\lim_{n \to +\infty} |u_n - \ell|^{s-r} = \frac{\lambda}{\mu}$$

avec $\lambda/\mu \neq 0$, $s - r > 0$ et $\lim_{n \to +\infty} (u_n - \ell) = 0$, ce qui est impossible.

▶ Si la convergence de $(u_n)_{n \in \mathbb{N}}$ vers ℓ est d'ordre $r \geq 2$, alors $\left|\frac{u_{n+1} - \ell}{u_n - \ell}\right|$ est équivalent à $\lambda |u_n - \ell|^{r-1}$ qui converge vers 0, c'est-à-dire que la convergence est rapide.

▶ Mais réciproquement une convergence rapide n'est pas obligatoirement super-linéaire comme le montre l'exemple de la suite $(u_n)_{n \in \mathbb{N}}$ définie par :

$$u_n = \sum_{k=0}^{n} \frac{1}{k!}$$

En effet, on a vu (exercice 6.4) que cette suite converge rapidement vers e avec $e - u_n \underset{+\infty}{\sim} \frac{1}{(n+1)!}$, de sorte que pour tout entier $r \geq 2$, on a :

$$\frac{e - u_{n+1}}{|e - u_n|^r} \underset{+\infty}{\sim} \frac{((n+1)!)^r}{(n+2)!} = \frac{((n+1)!)^{r-1}}{(n+2)} \underset{n \to +\infty}{\to} +\infty$$

et la convergence ne peut être d'ordre r.

Exercice 6.8 *Soient $I = [a, b]$ un intervalle réel fermé non réduit à un point et $f : I \to I$ de classe $\mathcal{C}^r$ avec $r \geq 2$ telle que $|f'(x)| < 1$ pour tout $x \in I$.*

1. *Montrer que f admet un unique point fixe $\ell \in I$.*
2. *On suppose que $f^{(k)}(\ell) = 0$ pour tout k compris ente 1 et $r - 1$ et $f^{(r)}(x) \neq 0$ pour tout $x \in I$. Pour u_0 donné dans $I \setminus \{\ell\}$, on définit la suite $u = (u_n)_{n\in\mathbb{N}}$ par $u_{n+1} = f(u_n)$. Montrer que cette suite converge vers ℓ et que la convergence est d'ordre r. On dit, dans cette situation, que ℓ est un point fixe super-attractif de f.*

Solution 6.8 *1. Voir l'exercice 6.7.*

2. La formule de Taylor-Lagrange à l'ordre r permet d'écrire que :

$$u_{n+1} - \ell = f(u_n) - f(\ell) = (u_n - \ell)^r \frac{f^{(r)}(c_n)}{r!}$$

avec c_n strictement compris entre u_n et ℓ, ce qui entraîne $u_n \neq \ell$ pour tout $n \geq 0$ (par récurrence) et :

$$\lim_{n\to+\infty} \frac{u_{n+1} - \ell}{(u_n - \ell)^r} = \frac{f^{(r)}(\ell)}{r!} \neq 0$$

ce qui implique que la suite u converge vers ℓ puisque :

$$\lim_{n\to+\infty} \frac{|u_{n+1} - \ell|}{|u_n - \ell|} = \lim_{n\to+\infty} |u_n - \ell|^{r-1} \frac{\left|f^{(r)}(c_n)\right|}{r!} = 0$$

et la convergence est d'ordre r.

Exercice 6.9 *On se donne une fonction g de classe $\mathcal{C}^3$ sur un intervalle fermé $I = [a, b]$ telle que $g(a)g(b) < 0$, $g'(x) \neq 0$ et $g''(x) \neq 0$ pour tout $x \in I$.*

1. *Montrer que l'équation $g(x) = 0$ admet une unique solution $\ell \in]a, b[$.*
2. *On note f la fonction définie sur I par $f(x) = x - \dfrac{g(x)}{g'(x)}$. Montrer que $f'(\ell) = 0$ et $f''(\ell) \neq 0$.*
3. *On désigne par $J = [\ell - \eta, \ell + \eta]$ un intervalle inclus dans I, avec $\eta > 0$, tel que $f''(x) \neq 0$ pour tout $x \in J$. Montrer que la suite $u = (u_n)_{n\in\mathbb{N}}$ définie par $u_0 \in J \setminus \{\ell\}$ et $u_{n+1} = u_n - \dfrac{g(u_n)}{g'(u_n)}$ pour tout $n \geq 0$ (méthode de Newton) converge vers ℓ et que la convergence est d'ordre 2.*

Solution 6.9 *1. Le théorème des valeurs intermédiaires nous dit que l'équation* $g(x) = 0$ *a au moins une solution dans* I *et l'hypothèse* $g'(x) \neq 0$ *pour tout* $x \in I$ *avec* g' *continue nous dit que* $g' > 0$ *ou* $g' < 0$ *(théorème des valeurs intermédiaires pour* g'*) ce qui implique que* g *est strictement monotone, donc injective. Cette solution* ℓ *est donc unique.*

2. Le réel ℓ *est l'unique point fixe dans* I *de la fonction* f *définie sur* I *par* $f(x) = x - \dfrac{g(x)}{g'(x)}$*. Cette fonction est de classe* $\mathcal{C}^2$ *sur* I *avec*

$$f'(x) = \frac{g(x)\, g''(x)}{(g'(x))^2}.$$

De $g(\ell) = 0$ *on déduit que* $f'(\ell) = 0$ *et :*

$$f''(\ell) = \lim_{x \to \ell} \frac{f'(x)}{x - \ell} = \lim_{x \to \ell} \frac{g(x)}{x - \ell} \frac{g''(x)}{(g'(x))^2} = \frac{g''(\ell)}{g'(\ell)} \neq 0.$$

3. Par continuité de f'' *il est possible de trouver un voisinage* J *de* ℓ *tel que* $f''(x) \neq 0$ *pour tout* $x \in J$*. La formule de Taylor-Lagrange à l'ordre* 2 *permet d'écrire que :*

$$u_{n+1} - \ell = f(u_n) - f(\ell) = (u_n - \ell)^2 \frac{f''(c_n)}{2!}$$

avec c_n *strictement compris entre* u_n *et* ℓ*, et* $f''(c_n) \neq 0$*, ce qui entraîne* $u_n \neq \ell$ *pour tout* $n \geq 0$ *(par récurrence) et :*

$$\lim_{n \to +\infty} \frac{u_{n+1} - \ell}{(u_n - \ell)^2} = \frac{f''(\ell)}{2} = \frac{g''(\ell)}{2g'(\ell)} \neq 0,$$

ce qui implique que la suite u *converge vers* ℓ *puisque*

$$\lim_{n \to +\infty} \frac{|u_{n+1} - \ell|}{|u_n - \ell|} = \lim_{n \to +\infty} |u_n - \ell| \frac{|f''(c_n)|}{2} = 0,$$

et que la convergence est d'ordre 2.

6.2 Accélération de la convergence

Définition 6.5 *Si* $v = (v_n)_{n \in \mathbb{N}}$ *est une autre suite convergente vers* ℓ *avec* $v_n \neq \ell$ *pour tout* $n \in \mathbb{N}$*, on dit que la convergence de cette suite vers* ℓ *est plus rapide que celle de* $u = (u_n)_{n \in \mathbb{N}}$ *si* $\displaystyle\lim_{n \to +\infty} \frac{v_n - \ell}{u_n - \ell} = 0$.

Accélérer la convergence d'une suite consiste à construire à partir de cette dernière une autre suite qui converge plus vite vers la même limite.

Exercice 6.10 *Soit* $(u_n)_{n\in\mathbb{N}^*}$ *la suite définie par :*

$$\forall n \geq 1,\ u_n = \left(1 + \frac{1}{n}\right)^n.$$

Montrer que la suite $(v_n)_{n\geq 0} = (u_{2^n})_{n\geq 0}$ *permet d'accélérer la convergence de cette suite.*

Solution 6.10 *On a vu (exercice 6.3) que* $e - u_n \underset{+\infty}{\backsim} \frac{e}{2n}$ *(convergence lente) et* $e - v_n \underset{+\infty}{\backsim} \frac{e}{2^{n+1}}$ *(convergence géométrique) de sorte que :*

$$\lim_{n\to+\infty} \frac{v_n - e}{u_n - e} = \lim_{n\to+\infty} \frac{n}{2^{n+1}} = 0.$$

De manière plus générale, un développement asymptotique de la forme :

$$u_n = \ell + \frac{\beta}{n^b} + o\left(\frac{1}{n^b}\right)$$

avec β non nul et $b > 0$ donne :

$$v_n = u_{2^n} = \ell + \frac{\beta}{2^{nb}} + o\left(\frac{1}{2^{nb}}\right)$$

et $\displaystyle\lim_{n\to+\infty} \frac{v_n - \ell}{u_n - \ell} = \lim_{n\to+\infty} \frac{n^b}{2^{nb}} = 0.$

Mais l'extraction ne permet pas toujours d'accélérer la convergence d'une suite. Par exemple pour $u = \left(\dfrac{1}{\ln(n)}\right)_{n\geq 2}$ et $v = \left(\dfrac{1}{\ln(n^2)}\right)_{n\geq 2}$, on a

$$\lim_{n\to+\infty} \frac{v_n}{u_n} = \frac{1}{\ln(2)}.$$

L'exercice qui suit nous donne un procédé élémentaire, mais peu performant, d'accélération de la convergence, l'idée étant de remplacer chaque u_n par une moyenne pondérée de u_n et u_{n+1}. Ce procédé sera affiné par la méthode de Richardson.

Exercice 6.11 *Soit* λ *un réel différent de* 1 *et* $v = (v_n)_{n\in\mathbb{N}}$ *la suite définie par :*

$$\forall n \geq 0,\ v_n = \frac{u_{n+1} - \lambda u_n}{1 - \lambda}$$

(v_n *est un barycentre de* u_{n+1} *et* u_n*).*

1. *Préciser pour quelles valeurs de λ et à quelles conditions sur u la suite v converge vers ℓ plus rapidement que u.*
2. *Appliquer ce procédé à la suite* $u = \left(\left(1+\frac{1}{2^n}\right)^{2^n}\right)_{n\in\mathbb{N}}$ *en précisant la vitesse de convergence de la suite accélératrice v obtenue.*

Solution 6.11 *Pour $\lambda = 0$, on a $v_n = u_{n+1}$, ce qui n'a pas beaucoup d'intérêt.*

1. *Pour tout réel λ différent de 1, la suite v converge vers ℓ et :*

$$\frac{v_n-\ell}{u_n-\ell} = \frac{u_{n+1}-\ell-\lambda\,(u_n-\ell)}{(1-\lambda)\,(u_n-\ell)} = \frac{1}{1-\lambda}\frac{u_{n+1}-\ell}{u_n-\ell} - \frac{\lambda}{1-\lambda},$$

de sorte que :

$$\lim_{n\to+\infty}\frac{v_n-\ell}{u_n-\ell} = 0 \quad\Leftrightarrow\quad \lim_{n\to+\infty}\frac{u_{n+1}-\ell}{u_n-\ell} = \lambda.$$

On a vu que la convergence de u vers ℓ impose $\lambda \in [-1,1]$ et comme $\lambda \neq 1$, on déduit que la suite v converge vers ℓ plus rapidement que u si, et seulement si,

$$\lim_{n\to+\infty}\frac{u_{n+1}-\ell}{u_n-\ell} = \lambda \in [-1,1[\,.$$

• *Pour $\lambda = -1$, on a $v_n = \dfrac{u_n+u_{n+1}}{2}$ et cette suite accélère u si, et seulement si, u converge lentement avec* $\lim\limits_{n\to+\infty}\dfrac{u_{n+1}-\ell}{u_n-\ell} = -1.$

• *Pour $0 < |\lambda| < 1$, la suite v accélère u si, et seulement si, la convergence de u est géométrique de rapport $|\lambda|$ avec* $\lim\limits_{n\to+\infty}\dfrac{u_{n+1}-\ell}{u_n-\ell} = \lambda.$

2. *On a vu que la convergence de la suite u vers $e \approx 2.718281828$ est géométrique de rapport $\lambda = 1/2$ (exercice 6.3), ce qui donne la suite v définie par :*

$$v_n = 2u_{n+1} - u_n.$$

Le développement limité à l'ordre 2 de $\exp\left(\dfrac{\ln(1+x)}{x}\right)$ *au voisinage de 0 nous donne le développement asymptotique :*

$$\left(1+\frac{1}{n}\right)^n = e\left(1-\frac{1}{2n}+\frac{11}{24}\frac{1}{n^2}+o\left(\frac{1}{n^2}\right)\right)$$

et :

$$u_n = e\left(1-\frac{1}{2^{n+1}}+\frac{11}{24}\frac{1}{2^{2n}}+o\left(\frac{1}{2^{2n}}\right)\right)$$

puis :

$$v_n = e\left(1 - \frac{11}{48}\frac{1}{2^{2n}} + o\left(\frac{1}{2^{2n}}\right)\right).$$

On a donc

$$e - v_n \underset{+\infty}{\backsim} \frac{11}{48}\frac{1}{2^{2n}}$$

et la convergence est géométrique de rapport $1/4$. *Pour* $n = 1$, *on a* $u_1 = 2.25$, $v_1 \approx 2.63281250$ *et pour* $n = 6$, $u_6 \approx 2.697344953$, $v_6 \approx 2.718133087$.

Exercice 6.12 *Considérons l'approximation du nombre* π *par la méthode d'Archimède des polygones réguliers. Cette méthode consiste à introduire la suite* $(u_n)_{n\geq 1}$ *définie par :*

$$\forall n \geq 1,\ u_n = 2^n \sin\left(\frac{\pi}{2^n}\right).$$

1. *Montrer que la suite* $(u_n)_{n\geq 1}$ *est aussi définie par :*

$$\begin{cases} u_1 = 2, \\ \forall n \geq 1,\ u_{n+1} = \sqrt{2}2^n\sqrt{1 - \sqrt{1 - \left(\frac{u_n}{2^n}\right)^2}}. \end{cases}$$

   Cette relation de récurrence permet un calcul itératif des u_n *sans utiliser le nombre* π *que l'on veut approcher.*

2. *Montrer que la convergence de cette suite vers* π *est géométrique de raison* $1/4$.
3. *Utiliser le procédé décrit à l'exercice précédent pour accélérer la convergence de cette suite en précisant la vitesse de convergence de la suite accélératrice* v *obtenue.*

Solution 6.12 *1. Avec :*

$$\sin^2\left(\frac{\pi}{2^{n+1}}\right) = \frac{1 - \cos\left(\frac{\pi}{2^n}\right)}{2} = \frac{1 - \sqrt{1 - \sin^2\left(\frac{\pi}{2^n}\right)}}{2},$$

on déduit que :

$$\forall n \geq 1,\ u_{n+1} = \sqrt{2}2^n\sqrt{1 - \sqrt{1 - \left(\frac{u_n}{2^n}\right)^2}}.$$

2. En utilisant le développement limité :

$$\sin(x) = x - \frac{x^3}{3!} + o\left(x^3\right),$$

on obtient le développement asymptotique :

$$u_n = \pi - \frac{\pi^3}{3!} \frac{1}{2^{2n}} + o\left(\frac{1}{2^{2n}}\right)$$

et $\pi - u_n \underset{+\infty}{\backsim} \frac{\pi^3}{3!} \frac{1}{2^{2n}}$. *Cette suite converge donc vers* π *et la convergence est géométrique de raison* $1/4$.

3. En utilisant le procédé décrit à l'exercice précédent, on accélère la convergence de cette suite en posant :

$$\forall n \geq 1,\ v_n = \frac{4u_{n+1} - u_n}{3}.$$

En poussant le développement limité précédent un peu plus loin, on a :

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} + o\left(x^5\right)$$

et le développement asymptotique :

$$u_n = \pi - \frac{\pi^3}{3!} \frac{1}{2^{2n}} + \frac{\pi^5}{5!} \frac{1}{2^{4n}} + o\left(\frac{1}{2^{4n}}\right)$$

qui donne :

$$v_n = \pi - \frac{\pi^5}{5!} \frac{3}{4} \frac{1}{2^{4n}} + o\left(\frac{1}{2^{4n}}\right),$$

soit

$$\pi - v_n \underset{+\infty}{\backsim} \frac{\pi^5}{5!} \frac{3}{4} \frac{1}{2^{4n}}.$$

Cette suite converge donc vers π *et la convergence est géométrique de raison* $1/16$.

Dans le cas où on dispose d'un encadrement de la forme :

$$\varepsilon'_n \leq u_n - \ell \leq \varepsilon_n$$

où $(\varepsilon'_n)_{n\in\mathbb{N}}$ et $(\varepsilon_n)_{n\in\mathbb{N}}$ sont des suites convergentes vers 0, la suite $(v_n)_{n\in\mathbb{N}}$ définie par $v_n = u_n - \varepsilon'_n$ converge aussi vers ℓ avec $0 \leq v_n - \ell \leq \delta_n = \varepsilon_n - \varepsilon'_n$, ce qui fournit parfois une suite accélératrice. Les exercices qui suivent nous fournissent des exemples de telle situation.

Exercice 6.13 *En utilisant l'encadrement obtenu avec l'exercice 6.4, montrer que la suite* $(v_n)_{n\geq 1}$ *définie par :*

$$\forall n \geq 1,\ v_n = \sum_{k=0}^{n} \frac{1}{k!} + \frac{1}{n \cdot n!},$$

accélère la convergence de la suite $(u_n)_{n\geq 1}$ *définie par :*

$$\forall n \geq 1,\ u_n = \sum_{k=0}^{n} \frac{1}{k!}.$$

Solution 6.13 *On a obtenu (exercice 6.4) l'encadrement :*

$$\frac{1}{(n+1)!} < e - u_n < \frac{1}{n \cdot n!},$$

qui donne :

$$0 < v_n - e < \frac{1}{n \cdot n!} - \frac{1}{(n+1)!} = \frac{1}{n \cdot (n+1)!}$$

et :

$$0 < \frac{v_n - e}{e - u_n} \leq \frac{1}{n} \underset{n\to+\infty}{\to} 0.$$

On a par exemple :

$$u_5 = \sum_{k=0}^{5} \frac{1}{k!} \approx 2.7167, \quad v_5 = u_5 + \frac{1}{5 * 5!} \approx 2.7183.$$

Exercice 6.14 *En utilisant l'encadrement obtenu avec l'exercice 6.5, montrer que la suite* $(v_n)_{n\geq 1}$ *définie par :*

$$\forall n \geq 1,\ v_n = \sum_{k=1}^{n} \frac{1}{k^\alpha} + \frac{1}{\alpha - 1} \frac{1}{(n+1)^{\alpha-1}},$$

accélère la convergence de la suite $(u_n)_{n\geq 1}$ *définie par :*

$$\forall n \geq 1,\ u_n = \sum_{k=1}^{n} \frac{1}{k^\alpha}$$

Solution 6.14 *On a obtenu (exercice 6.5) l'encadrement :*

$$\forall n \geq 1,\ \frac{1}{(n+1)^{\alpha-1}} \leq (\alpha - 1)(\ell - u_n) \leq \frac{1}{n^{\alpha-1}},$$

qui donne :

$$\forall n \geq 1, \quad 0 \leq \ell - v_n \leq \frac{1}{\alpha - 1} \frac{(n+1)^{\alpha-1} - n^{\alpha-1}}{n^{\alpha-1} (n+1)^{\alpha-1}},$$

et :

$$0 \leq \frac{\ell - v_n}{\ell - u_n} \leq \frac{(n+1)^{\alpha-1} - n^{\alpha-1}}{n^{\alpha-1}} \underset{+\infty}{\backsim} \frac{\alpha - 1}{n} \underset{n \to +\infty}{\to} 0.$$

Exercice 6.15 *En utilisant l'encadrement obtenu avec l'exercice 6.6, montrer que la suite $(v_n)_{n \geq 1}$ définie par :*

$$\forall n \geq 1, \ v_n = \sum_{k=1}^{n} \frac{1}{k} - \ln(n+1) + \frac{1}{2(n+1)}$$

accélère la convergence de la suite $(u_n)_{n \geq 1}$ définie par :

$$\forall n \geq 1, \ u_n = \sum_{k=1}^{n} \frac{1}{k} - \ln(n+1)$$

Solution 6.15 *On a obtenu (exercice 6.6) l'encadrement :*

$$\frac{1}{2(n+1)} < \gamma - u_n < \frac{1}{2n}$$

qui donne :

$$0 < \gamma - v_n < \frac{1}{2n(n+1)}$$

et :

$$0 \leq \frac{\gamma - v_n}{\gamma - u_n} \leq \frac{1}{n} \underset{n \to +\infty}{\to} 0.$$

6.3 Méthode d'accélération d'Aitken

Pour ce paragraphe, on suppose que suite $(u_n)_{n \in \mathbb{N}}$ converge vers un réel ℓ (toujours avec $u_n \neq \ell$ pour tout $n \in \mathbb{N}$) avec :

$$\lim_{n \to +\infty} \frac{u_{n+1} - \ell}{u_n - \ell} = \lambda \in]-1, 1[\setminus \{0\}$$

c'est-à-dire que la convergence est géométrique de rapport $|\lambda|$. On suppose de plus que $u_{n+1} \neq u_n$ pour tout $n \in \mathbb{N}$.

On a vu avec l'exercice 6.11 que la suite $v = (v_n)_{n\in\mathbb{N}}$ des moyennes pondérées définie par :

$$v_n = \frac{u_{n+1} - \lambda u_n}{1-\lambda}$$

est une suite accélératrice de u.
Mais dans la pratique le coefficient λ peut être prévu sans connaître explicitement sa valeur, de sorte que ce procédé d'accélération n'est pas utilisable directement.
Un exemple typique d'une telle situation est fourni par une suite d'approximations successives d'un point fixe attractif ℓ d'une fonction f de classe $\mathcal{C}^1$, le coefficient $\lambda = f'(\ell)$ qui fait intervenir le réel ℓ qu'on cherche à approximer, étant inconnu dans $]-1,1[\setminus\{0\}$ (voir l'exercice 6.7).
La méthode d'Aitken nous permet de construire une suite de réels $(\lambda_n)_{n\in\mathbb{N}}$ qui va converger vers λ et on définira une suite accélératrice par les moyennes pondérées :

$$v_n = \frac{u_{n+1} - \lambda_n u_n}{1-\lambda_n}.$$

On rappelle que la suite $(\lambda_n)_{n\in\mathbb{N}^*}$ définie par :

$$\forall n \in \mathbb{N}^*,\ \lambda_n = \frac{u_{n+1} - u_n}{u_n - u_{n-1}}$$

converge vers λ (lemme 6.1). Comme $\lambda < 1$, il existe un entier n_0 tel que l'on ait $\lambda_n < 1$ pour tout $n \geq n_0$. On peut donc définir la suite $(v_n)_{n\geq n_0}$ par :

$$\forall n \geq n_0,\ v_n = \frac{u_{n+1} - \lambda_n u_n}{1-\lambda_n}.$$

Théorème 6.1 *La suite $(v_n)_{n\geq n_0}$ converge vers ℓ plus rapidement que la suite $(u_n)_{n\in\mathbb{N}}$.*

Preuve. Pour tout $n \geq n_0$, on a :

$$\begin{aligned}\frac{v_n-\ell}{u_n-\ell} = \frac{\dfrac{u_{n+1}-\lambda_n u_n}{1-\lambda_n} - \ell}{u_n-\ell} &= \frac{1}{1-\lambda_n}\frac{u_{n+1}-\lambda_n u_n - \ell(1-\lambda_n)}{u_n-\ell} \\ &= \frac{1}{1-\lambda_n}\left(\frac{u_{n+1}-\ell}{u_n-\ell} - \lambda_n\right) \underset{n\to+\infty}{\to} 0.\end{aligned}$$

■

En utilisant la définition des λ_n, on peut écrire chaque terme de la suite accélératrice de Aitken $(v_n)_{n\geq n_0}$ sous la forme :

$$v_n = \frac{u_{n+1}u_{n-1} - u_n^2}{u_{n+1} - 2u_n + u_{n-1}} = u_{n-1} - \frac{(u_n - u_{n-1})^2}{u_{n+1} - 2u_n + u_{n-1}}$$

(comme $\lambda_n = \dfrac{u_{n+1} - u_n}{u_n - u_{n-1}} < 1$, on a $(u_{n+1} - u_n) - (u_n - u_{n-1}) \neq 0$ et le dénominateur de v_n est bien non nul).
En introduisant les opérateurs de Aitken Δ et Δ^2 définis par :

$$\begin{cases} \Delta u_n = u_{n+1} - u_n \\ \Delta^2 u_n = \Delta(\Delta u_n) = u_{n+2} - 2u_{n+1} + u_n \end{cases}$$

on a :

$$v_{n+1} = u_n - \frac{(\Delta u_n)^2}{\Delta^2 u_n}.$$

Exercice 6.16 *Soit $(u_n)_{n\in\mathbb{N}}$ une suite arithmético-géométrique définie par $u_0 \in \mathbb{R}$ et $u_{n+1} = au_n + b$ où a, b sont des réels donnés avec $0 < |a| < 1$.*

1. *Montrer que u converge ver $\ell = \dfrac{b}{1-a}$ et préciser sa vitesse de convergence dans le cas où $u_0 \neq \ell$.*
2. *Décrire la suite accélératrice de Aitken correspondante.*

Solution 6.16 *1. Si cette suite converge, alors sa limite est solution de l'équation $x = ax + b$ qui a pour unique solution $\ell = \dfrac{b}{1-a}$. De $u_{n+1} = au_n + b$ et $\ell = a\ell + b$, on déduit que $u_{n+1} - \ell = a(u_n - \ell)$ et par récurrence*

$$u_n - \ell = a^n (u_0 - \ell),$$

soit $u_n = \ell + a^n(u_0 - \ell)$ et $\lim_{n\to+\infty} u_n = \ell$ avec $u_n \neq \ell$ pour tout $n \geq 0$ et

$$\frac{u_{n+1} - \ell}{u_n - \ell} = a,$$

c'est-à-dire que la convergence est géométrique de rapport $|a|$ (ce qui n'est pas étonnant).

2. On a $u_{n+1} - u_n = a^n(u_0 - \ell)(a-1)$ et $u_{n+1} - au_n = b = \ell(1-a)$, de sorte que :

$$\lambda_n = \frac{u_{n+1} - u_n}{u_n - u_{n-1}} = a \text{ et } v_n = \frac{u_{n+1} - \lambda_n u_n}{1 - \lambda_n} = \ell.$$

La suite v est donc stationnaire sur ℓ.

Exercice 6.17 *On reprend la situation de l'exercice 6.7 avec $I = [a, b]$ et $f : I \to I$ de classe $\mathcal{C}^2$ telle que $0 < |f'(x)| < 1$ pour tout $x \in I$.*
On a vu que cette fonction admet un unique point fixe (attractif) $\ell \in I$ et que la suite $u = (u_n)_{n\in\mathbb{N}}$ définie par $u_0 \in I \setminus \{\ell\}$ et $u_{n+1} = f(u_n)$ pour tout $n \geq 0$ converge vers ℓ avec $u_n \neq \ell$ pour tout $n \in \mathbb{N}$, la convergence étant géométrique de rapport $|f'(\ell)|$.
Montrer que, si $f''(\ell) \neq 0$, alors la convergence de la suite accélératrice v de Aitken correspondante est géométrique de rapport $(f'(\ell))^2$.

Solution 6.17 *Un développement limité à l'ordre 2 en ℓ nous donne, en tenant compte de $f(\ell) = \ell$:*

$$u_n - \ell = f'(\ell)(u_{n-1} - \ell) + \frac{f''(\ell)}{2}(u_{n-1} - \ell)^2 + o((u_{n-1} - \ell)^2).$$

En notant $e_n = u_n - \ell$, $\lambda = f'(\ell)$ et $\mu = \frac{f''(\ell)}{2}$, cela s'écrit :

$$e_n = \lambda e_{n-1} + \mu e_{n-1}^2 + o\left(e_{n-1}^2\right)$$

et, en désignant par $(v_n)_{n\geq n_0}$ la suite accélératrice de Aitken, on a :

$$v_n - \ell = u_{n-1} - \ell - \frac{(\Delta u_{n-1})^2}{\Delta^2 u_{n-1}}$$

avec :

$$\begin{cases} \Delta u_{n-1} = u_n - u_{n-1} = e_n - e_{n-1} \\ \Delta^2 u_{n-1} = u_{n+1} - 2u_n + u_{n-1} = e_{n+1} - 2e_n + e_{n-1} \end{cases}$$

ce qui donne :

$$v_n - \ell = e_{n-1} - \frac{(e_n - e_{n-1})^2}{e_{n+1} - 2e_n + e_{n-1}}$$

avec :

$$\begin{cases} e_n - e_{n-1} = (\lambda - 1)e_{n-1} + \mu e_{n-1}^2 + o\left(e_{n-1}^2\right), \\ (e_n - e_{n-1})^2 = (\lambda - 1)^2 e_{n-1}^2 + 2\mu(\lambda - 1)e_{n-1}^3 + o\left(e_{n-1}^3\right), \\ e_{n+1} - e_n = \lambda(\lambda - 1)e_{n-1} + \left((\lambda - 1)\mu + \lambda^2\mu\right)e_{n-1}^2 + o\left(e_{n-1}^2\right), \\ \begin{aligned} e_{n+1} - 2e_n + e_{n-1} &= (e_{n+1} - e_n) - (e_n - e_{n-1}) \\ &= (\lambda - 1)^2 e_{n-1} + (\lambda - 1)(\lambda + 2)\mu e_{n-1}^2 + o\left(e_{n-1}^2\right) \end{aligned} \end{cases}$$

et donc :

$$\begin{aligned} v_n - \ell &= e_{n-1} - \frac{(\lambda - 1)e_{n-1} + 2\mu e_{n-1}^2 + o\left(e_{n-1}^2\right)}{(\lambda - 1) + \mu(\lambda + 2)e_{n-1} + o(e_{n-1})} \\ &= \frac{\lambda\mu + o(1)}{(\lambda - 1) + o(1)} e_{n-1}^2. \end{aligned}$$

ou encore :

$$\lim_{n\to+\infty} \frac{v_n - \ell}{e_{n-1}^2} = \frac{\lambda\mu}{\lambda - 1}.$$

Comme $f''(\ell) \neq 0$, *on a* $\mu \neq 0$ *et :*

$$v_n - \ell \underset{+\infty}{\backsim} \frac{\lambda\mu}{1-\lambda} e_{n-1}^2 = \frac{f'(\ell) f''(\ell)}{2(1 - f'(\ell))} e_{n-1}^2.$$

On a donc :

$$\frac{v_{n+1} - \ell}{v_n - \ell} \underset{+\infty}{\backsim} \left(\frac{e_n}{e_{n-1}}\right)^2 \underset{+\infty}{\backsim} \lambda^2 = \left(f'(\ell)\right)^2.$$

En définitive, on est passé d'une convergence géométrique de rapport $|f'(\ell)|$ *à une convergence géométrique de rapport* $(f'(\ell))^2$, *ce qui confirme l'accélération de la convergence puisque* $|f'(\ell)| < 1$.

Remarque 6.5 *Si, avec les notations de l'exercice précédent, on désigne pour tout entier naturel* n, *par* M_n *le point de* $\mathbb{R}^2$ *de coordonnées*

$$(u_n, f(u_n)) = (u_n, u_{n+1})$$

dans la base canonique, la pente de la droite $(M_n M_{n+1})$ *est :*

$$\delta_n = \frac{u_{n+2} - u_{n+1}}{u_{n+1} - u_n} = \frac{\Delta_{n+1}}{\Delta_n}$$

et l'équation de cette droite est :

$$y = u_{n+1} + \frac{\Delta u_{n+1}}{\Delta u_n}(x - u_n).$$

Le point d'intersection de cette droite avec la première bissectrice est le point M'_n *de coordonnées* (x_n, x_n) *où* x_n *est solution de :*

$$x = u_{n+1} + \frac{\Delta u_{n+1}}{\Delta u_n}(x - u_n) = \Delta u_n + u_n + \frac{\Delta u_{n+1}}{\Delta u_n}(x - u_n)$$

soit de :

$$(x - u_n)\left(1 - \frac{\Delta u_{n+1}}{\Delta u_n}\right) = \Delta u_n$$

ce qui donne :

$$\begin{aligned} x_n &= u_n + \frac{\Delta u_n}{1 - \frac{\Delta u_{n+1}}{\Delta u_n}} = u_n - \frac{(\Delta u_n)^2}{\Delta u_{n+1} - \Delta u_n} \\ &= u_n - \frac{(\Delta u_n)^2}{\Delta^2 u_n} = v_{n+1}. \end{aligned}$$

Remarque 6.6 *Pour la programmation de la méthode de Aitken on peut remarquer que les v_n s'écrivent aussi :*

$$v_n = u_n + \left(\frac{1}{u_{n+1} - u_n} - \frac{1}{u_n - u_{n-1}} \right)^{-1}.$$

En effet, on a :

$$\begin{aligned} v_n &= \frac{u_{n+1}u_{n-1} - u_n^2}{u_{n+1} - 2u_n + u_{n-1}} \\ &= u_n - \frac{(u_{n+1} - u_n)(u_n - u_{n-1})}{(u_{n+1} - u_n) - (u_n - u_{n-1})} \\ &= u_n + \frac{1}{\dfrac{1}{u_{n+1} - u_n} - \dfrac{1}{u_n - u_{n-1}}}. \end{aligned}$$

Dans le cas où la suite $(u_n)_{n\in\mathbb{N}}$ est une suite d'approximations successives définie par $u_{n+1} = f(u_n)$ (avec les notations et hypothèses de l'exercice précédent), en définissant la fonction g par

$$g(x) = \frac{1}{f(x) - x}$$

pour $x \in I \setminus \{\ell\}$, les v_n peuvent se calculer comme suit :

$$\begin{cases} u_n = f(u_{n-1}), \\ v_n = u_n + \dfrac{1}{g(u_n) - g(u_{n-1})}. \end{cases}$$

Exemple 6.1 *On s'intéresse aux points fixes de la fonction $f : x \mapsto e^{-x}$ sur $[0, +\infty[$. Cette fonction est indéfiniment dérivable, strictement décroissante sur $[0, +\infty[$ avec $f'(x) = -e^{-x}$ et $f([0, +\infty[) =]0, 1]$.*
En posant $I = [e^{-1}, 1] = [0.367\,88, 1]$, on a $f(I) = \left[e^{-1}, e^{-e^{-1}}\right] \subset I$ et $0 < |f'(x)| \leq \lambda = e^{-1} < 1$ pour tout $x \in I$. La suite $(u_n)_{n\in\mathbb{N}}$ définie par $u_0 = 0.5$ et $u_{n+1} = f(u_n)$ pour tout $n \geq 0$ converge donc vers l'unique point fixe $\ell \in I$, la convergence étant géométrique.
Le programme Maple qui suit nous donne les valeurs des u_n et v_n pour n compris entre 1 et 15 :

```
restart; f :=x->exp(-x); g :=x->1/(f(x)-x); i :=0; x :=0.5;
for n from 1 to 15 do
    i :=i+1; z :=f(x); y :=z+1/(g(z)-g(x)); x :=z;
od;
```

Ce qui donne les valeurs suivantes :

$$\begin{cases} u_1 = 0.6065306597 \\ v_1 = 0.5676238764 \end{cases} \begin{cases} u_{14} = 0.5671188642 \\ v_{14} = 0.5671432906 \end{cases} \begin{cases} u_{15} = 0.5671571437 \\ v_{15} = 0.5671432905 \end{cases}$$

On constate que la convergence de la suite $(v_n)_{n\in\mathbb{N}}$ *vers le point fixe de* f, $\ell = 0.567143290409$, *est plus rapide que celle de la suite* $(u_n)_{n\in\mathbb{N}}$.
En utilisant $f' = -f$, $f'' = f$ *et* $f(\ell) = \ell$, *on a ici :*

$$\begin{cases} v_n - \ell \underset{+\infty}{\backsim} \dfrac{f'(\ell)\, f''(\ell)}{2\,(1 - f'(\ell))} (u_{n-1} - \ell)^2 = \dfrac{\ell^2}{2\,(1+\ell)} (u_{n-1} - \ell)^2 \\ \dfrac{v_{n+1} - \ell}{v_n - \ell} \underset{+\infty}{\backsim} (f'(\ell))^2 = \ell^2 \end{cases}$$

avec :

$$\begin{cases} \ell^2 \approx 0.321651511855947 \\ \dfrac{\ell^2}{2\,(1+\ell)} \approx 0.102623516887215. \end{cases}$$

Dans le cas d'un point fixe super-attractif (c'est-à-dire que $f'(\ell) = 0$), en supposant de plus que $f''(\ell) \neq 0$, on a, pour f de classe $\mathcal{C}^4$, le développement asymptotique suivant de l'erreur d'approximation :

$$e_n = \lambda_2 e_{n-1}^2 + \lambda_3 e_{n-1}^3 + \lambda_4 e_{n-1}^4 + o\left(e_{n-1}^4\right),$$

où on a noté :

$$\lambda_2 = \frac{f''(\ell)}{2},\ \lambda_3 = \frac{f^{(3)}(\ell)}{3!},\ \lambda_4 = \frac{f^{(4)}(\ell)}{3!},$$

ce qui donne :

$$\begin{cases} (e_n - e_{n-1})^2 = e_{n-1}^2 - 2\lambda_2 e_{n-1}^3 + \left(\lambda_2^2 - 2\lambda_3\right) e_{n-1}^4 + o\left(e_{n-1}^4\right) \\ e_{n+1} - e_n \ = \ \lambda_2 e_n^2 + o\left(e_n^2\right) - e_n \\ \qquad\qquad = \ -\lambda_2 e_{n-1}^2 - \lambda_3 e_{n-1}^3 + \left(\lambda_2^3 - \lambda_4\right) e_{n-1}^4 + o\left(e_{n-1}^4\right) \\ e_{n+1} - 2e_n + e_{n-1} = e_{n-1} - 2\lambda_2 e_{n-1}^2 - 2\lambda_3 e_{n-1}^3 + \left(\lambda_2^3 - 2\lambda_4\right) e_{n-1}^4 + o\left(e_{n-1}^4\right) \end{cases}$$

et :

$$v_n - \ell = \frac{-\lambda_2^2 + o(1)}{1 - 2\lambda_2 e_{n-1} + o(e_{n-1})} e_{n-1}^3 = \left(-\lambda_2^2 + o(1)\right) e_{n-1}^3,$$

soit :

$$v_n - \ell \underset{+\infty}{\backsim} \frac{(f''(\ell))^2}{4} (u_n - \ell)^3.$$

Avec

$$\begin{cases} v_n - \ell = \left(-\lambda_2^2 + o(1)\right) e_{n-1}^3, \\ v_{n+1} - \ell = \left(-\lambda_2^2 + o(1)\right) e_n^3 = \left(-\lambda_2^5 + o(1)\right) e_{n-1}^6, \end{cases}$$

on déduit que :

$$\frac{v_{n+1} - \ell}{(v_n - \ell)^2} \backsim \frac{f''(\ell)}{2}.$$

La convergence vers ℓ de $(v_n)_{n \geq n_0}$ est donc, comme celle de $(u_n)_{n \in \mathbb{N}}$, d'ordre 2, mais cette convergence est quand même plus rapide.

Exercice 6.18 *Donner une expression de la suite accélératrice de Aitken pour la suite* $(u_n)_{n \geq 1}$ *définie par :*

$$\forall n \geq 1, \; u_n = \sum_{k=1}^{n} \frac{1}{k^\alpha}$$

où $\alpha > 1$. *Préciser la vitesse de convergence de cette suite accélératrice.*

Solution 6.18 *On a :*

$$\begin{cases} \Delta u_n = u_{n+1} - u_n = \dfrac{1}{(n+1)^\alpha} \\ \Delta^2 u_n = \Delta(\Delta u_n) = \dfrac{1}{(n+2)^\alpha} - \dfrac{1}{(n+1)^\alpha} \end{cases}$$

et :

$$\begin{aligned} v_{n+1} = u_n - \frac{(\Delta u_n)^2}{\Delta^2 u_n} &= u_n - \frac{(n+2)^\alpha}{(n+1)^{2\alpha} - (n+2)^\alpha (n+1)^\alpha} \\ &= u_n + \frac{(n+2)^\alpha}{(n+1)^\alpha \left((n+2)^\alpha - (n+1)^\alpha\right)}. \end{aligned}$$

Par exemple, pour $n = 2$, *on a :*

$$v_{n+1} = u_n + \frac{(n+2)^2}{(n+1)^2 (2n+3)}$$

En reprenant les calculs de l'exercice 6.5, on a pour tout $n \geq 1$:

$$\frac{1}{\alpha - 1} \frac{1}{(n+1)^{\alpha - 1}} \leq \ell - u_n \leq \frac{1}{\alpha - 1} \frac{1}{n^{\alpha - 1}}$$

et donc :

$$\begin{aligned}\tfrac{1}{(\alpha-1)(n+1)^{\alpha-1}}\left(1-\tfrac{(\alpha-1)(n+2)^{\alpha}}{(n+1)((n+2)^{\alpha}-(n+1)^{\alpha})}\right) &\leq \ell - v_{n+1}\\ &\leq \tfrac{1}{(\alpha-1)n^{\alpha-1}}\left(1-\tfrac{n^{\alpha-1}(\alpha-1)(n+2)^{\alpha}}{(n+1)^{\alpha}((n+2)^{\alpha}-(n+1)^{\alpha})}\right)\end{aligned}$$

soit :

$$\begin{aligned}\tfrac{1}{(\alpha-1)(n+1)^{\alpha-1}}\left(1-\tfrac{(\alpha-1)}{(n+1)\left(1-\left(\frac{n+1}{n+2}\right)^{\alpha}\right)}\right) &\leq \ell - v_{n+1}\\ &\leq \tfrac{1}{(\alpha-1)n^{\alpha-1}}\left(1-\tfrac{n^{\alpha-1}(\alpha-1)}{(n+1)^{\alpha}\left(1-\left(\frac{n+1}{n+2}\right)^{\alpha}\right)}\right)\end{aligned}$$

avec :

$$\begin{aligned}1-\left(\frac{n+1}{n+2}\right)^{\alpha} &= 1-\left(1-\frac{1}{n+2}\right)^{\alpha}\\ &= 1-\left(1-\alpha\frac{1}{n+2}+o\left(\frac{1}{n}\right)\right)\\ &= \alpha\frac{1}{n+2}+o\left(\frac{1}{n}\right) \underset{n\to+\infty}{\sim} \frac{\alpha}{n}\end{aligned}$$

ce qui donne :

$$\frac{n^{\alpha-1}(\alpha-1)}{(n+1)^{\alpha}\left(1-\left(\frac{n+1}{n+2}\right)^{\alpha}\right)} \underset{n\to+\infty}{\sim} \frac{n^{\alpha-1}(\alpha-1)}{n^{\alpha}\dfrac{\alpha}{n}} = \frac{\alpha-1}{\alpha}$$

et :

$$\frac{(\alpha-1)}{(n+1)\left(1-\left(\frac{n+1}{n+2}\right)^{\alpha}\right)} \underset{n\to+\infty}{\sim} \frac{\alpha-1}{n\dfrac{\alpha}{n}} = \frac{\alpha-1}{\alpha}$$

Il en résulte que :

$$\ell - v_{n+1} \underset{n\to+\infty}{\sim} \tfrac{1}{(\alpha-1)n^{\alpha-1}}\left(1-\frac{\alpha-1}{\alpha}\right) = \frac{1}{\alpha(\alpha-1)n^{\alpha-1}}$$

et la convergence de la suite $(v_n)_{n\geq 1}$ *est lente.*

6.4 Méthode d'accélération de Richardson

On se place dans un premier temps dans le cas où on dispose d'un développement asymptotique de la forme :

$$u_n = \ell + \beta\lambda^n + \gamma\mu^n + o(\mu^n)$$

avec β, γ non nuls et $0 < |\mu| < |\lambda| < 1$. La suite $(u_n)_{n\in\mathbb{N}}$ converge donc vers ℓ et la convergence est géométrique de rapport $|\lambda|$.
Si on connaît explicitement les coefficients β et λ, on peut accélérer la convergence de la suite $(u_n)_{n\in\mathbb{N}}$ en la remplaçant par la suite $(v_n)_{n\in\mathbb{N}}$ définie par :

$$v_n = u_n - \beta\lambda^n.$$

Cette suite converge bien vers ℓ et avec $v_n - \ell \underset{+\infty}{\backsim} \gamma\mu^n$, la convergence est donc géométrique de rapport $|\mu|$ et :

$$\frac{v_n - \ell}{u_n - \ell} \underset{+\infty}{\backsim} \frac{\gamma}{\beta}\left(\frac{\mu}{\lambda}\right)^n \underset{n\to+\infty}{\to} 0,$$

ce qui confirme bien que la suite $(v_n)_{n\in\mathbb{N}}$ converge vers ℓ plus vite que la suite $(u_n)_{n\in\mathbb{N}}$.
Si on connaît explicitement le coefficient λ, mais pas le coefficient β, on peut définir un barycentre de u_{n+1} et u_n où le terme $\beta\lambda^n$ a été éliminé. Pour ce faire, on écrit que :

$$\begin{cases} u_{n+1} = \ell + \beta\lambda^{n+1} + \mu^{n+1}(\gamma + o(1)) \\ u_n = \ell + \beta\lambda^n + \mu^n(\gamma + o(1)) \end{cases}$$

et :

$$u_{n+1} - \lambda u_n = (1-\lambda)\,\ell + \mu^n\left((\mu - \lambda)\,\gamma + o(1)\right),$$

ce qui nous conduit à introduire la suite $(v_n)_{n\in\mathbb{N}}$ définie par :

$$v_n = \frac{u_{n+1} - \lambda u_n}{1 - \lambda}.$$

C'est la situation décrite à l'exercice 6.11.
On a alors :

$$v_n - \ell = \mu^n\left(\frac{\mu - \lambda}{1 - \lambda}\gamma + o(1)\right) \underset{+\infty}{\backsim} \frac{\mu - \lambda}{1 - \lambda}\gamma\mu^n$$

c'est-à-dire que la convergence est géométrique de rapport $|\mu|$ et :

$$\frac{v_n - \ell}{u_n - \ell} \underset{+\infty}{\backsim} \frac{\mu - \lambda}{1 - \lambda}\frac{\gamma}{\beta}\left(\frac{\mu}{\lambda}\right)^n \underset{n\to+\infty}{\to} 0.$$

On est donc ainsi passé de la suite $(u_n)_{n\in\mathbb{N}}$ qui converge vers ℓ avec une vitesse de convergence géométrique de raison $|\lambda|$ à la suite $(v_n)_{n\in\mathbb{N}}$ qui converge aussi vers ℓ avec une vitesse de convergence géométrique de raison $|\mu| < |\lambda|$.
Les exercices 6.11 et 6.12 nous donnent des exemples de cette situation où on approxime respectivement les nombres e et π.

De manière plus générale, si on dispose d'un développement asymptotique de la forme :

$$u_n = \ell + \frac{\beta}{n} + \frac{\gamma}{n^2} + o\left(\frac{1}{n^2}\right)$$

avec β et γ non nuls, on utilise la suite $(v_n)_{n\in\mathbb{N}} = (x_{2^n})_{n\in\mathbb{N}}$ pour laquelle on a le développement asymptotique :

$$v_n = \ell + \frac{\beta}{2^n} + \frac{\gamma}{4^n} + o\left(\frac{1}{4^n}\right)$$

et une suite accélératrice est définie par :

$$w_n = 2v_{n+1} - v_n.$$

On a alors :

$$\begin{cases} v_n - \ell \underset{+\infty}{\backsim} \dfrac{\beta}{2^n} \\ w_n - \ell \underset{+\infty}{\backsim} -\dfrac{1}{2}\dfrac{\gamma}{4^n} \end{cases}$$

c'est-à-dire qu'on passe d'une convergence géométrique de raison 1/2 à une convergence géométrique de raison 1/4 (voir l'exercice 6.11).

La méthode de Richardson consiste à itérer le procédé précédent dès que l'on dispose d'un développement asymptotique de la forme :

$$u_n = \ell + \sum_{j=1}^{p+1} \beta_j \lambda_j^n + o\left(\lambda_{p+1}^n\right)$$

où p est un entier naturel non nul, les coefficients β_j sont tous non nuls et les coefficients λ_j tels que :

$$0 < |\lambda_{p+1}| < |\lambda_p| < \cdots < |\lambda_1| < 1.$$

Si tous les coefficients β_j et λ_j sont connus, on peut accélérer la convergence de cette suite en introduisant la suite $(v_n)_{n\in\mathbb{N}}$ définie par :

$$v_n = u_n - \sum_{j=1}^{p+1} \beta_j \lambda_j^n.$$

Ce cas se présente pour les sommes de séries numériques de la forme $\sum_{n=0}^{+\infty} f(n)$, où la fonction f est indéfiniment dérivable sur $\mathbb{R}^{+,*}$. Le développement asymptotique est obtenu en utilisant la formule d'Euler et Mac-Laurin en supposant

le calcul explicite des dérivées de la fonction f facilement réalisable. C'est le cas par exemple pour les séries de Riemann convergentes.
Si les coefficients λ_j sont tous connus, mais pas les coefficients β_j, on va les éliminer progressivement en itérant le procédé barycentrique décrit précédemment, ce qui nous amène à introduire, pour tout entier k compris entre 0 et p, les suites $(u_{n,k})_{n\in\mathbb{N}}$ définies par les formules de récurrence suivantes :

$$\begin{cases} u_{n,0} = u_n, \\ u_{n,k} = \dfrac{u_{n+1,k-1} - \lambda_k u_{n,k-1}}{1-\lambda_k}. \end{cases}$$

Lemme 6.2 *Avec les notations et hypothèses qui précèdent, on a pour tout entier k compris entre 0 et p, le développement asymptotique :*

$$u_{n,k} = \ell + \sum_{j=k+1}^{p+1} \beta_{k,j}\lambda_j^n + o\left(\lambda_{p+1}^n\right),$$

les coefficients $\beta_{k,j}$ étant tous non nuls.

Preuve. On procède par récurrence finie sur k.Pour $k = 0$ c'est l'hypothèse. En supposant le résultat acquis au rang $k - 1 < p$, on a :

$$\begin{cases} u_{n+1,k-1} = \ell + \sum\limits_{j=k}^{p+1} \beta_{k-1,j}\lambda_j^{n+1} + o\left(\lambda_{p+1}^{n+1}\right) \\ u_{n,k-1} = \ell + \sum\limits_{j=k}^{p+1} \beta_{k-1,j}\lambda_j^{n} + o\left(\lambda_{p+1}^{n}\right) \end{cases}$$

l'élimination du coefficient $\beta_{k-1,k}$ entre ces deux équations se faisant avec :

$$\frac{u_{n+1,k-1} - \lambda_k u_{n,k-1}}{1-\lambda_k} = \ell + \sum_{j=k+1}^{p+1} \beta_{k,j}\lambda_j^n + o\left(\lambda_{p+1}^n\right)$$

où $\beta_{k,j} = \dfrac{\lambda_j - \lambda_k}{1-\lambda_k}\beta_{k-1,j} \neq 0$ pour tout j compris entre $k+1$ et $p+1$. ■

Théorème 6.2 (Richardson) *Avec les notations et hypothèses qui précèdent, pour tout entier k compris entre 1 et p, la suite $(u_{n,k})_{n\in\mathbb{N}}$ converge vers ℓ plus rapidement que la suite $(u_{n,k-1})_{n\in\mathbb{N}}$, la convergence de la suite $(u_{n,k})_{n\in\mathbb{N}}$ étant géométrique de raison $|\lambda_{k+1}|$. Plus précisément, pour tout k compris entre 0 et p, on a :*

$$u_{n,k} - \ell \underset{+\infty}{\backsim} \beta_{k,k+1}\lambda_{k+1}^n$$

avec :

$$\beta_{k,k+1} = \beta_{k+1} \prod_{j=1}^{k} \frac{\lambda_{k+1} - \lambda_j}{1 - \lambda_j}.$$

Preuve. Avec l'hypothèse $0 < |\lambda_{p+1}| < \cdots < |\lambda_{k+1}| < |\lambda_k|$ et lemme précédent, on déduit que, pour k compris entre 1 et p, on a :

$$\begin{cases} u_{n,k} - \ell \underset{+\infty}{\backsim} \beta_{k,k+1}\lambda_{k+1}^n \\ \dfrac{u_{n,k} - \ell}{u_{n,k-1} - \ell} \underset{+\infty}{\backsim} \dfrac{\beta_{k,k+1}}{\beta_{k-1,k}} \left(\dfrac{\lambda_{k+1}}{\lambda_k}\right)^n \underset{n\to+\infty}{\to} 0 \end{cases}$$

avec :

$$\beta_{k,k+1} = \frac{\lambda_{k+1} - \lambda_k}{1 - \lambda_k} \frac{\lambda_{k+1} - \lambda_{k-1}}{1 - \lambda_{k-1}} \cdots \frac{\lambda_{k+1} - \lambda_1}{1 - \lambda_1} \beta_{k+1}$$

ce qui est le résultat annoncé. ■

Exercice 6.19 *On reprend l'exemple de la suite d'Archimède $u = (u_n)_{n\geq 1}$ permettant d'approcher le nombre π. On rappelle qu'elle est définie par :*

$$\forall n \geq 1,\ u_n = 2^n \sin\left(\frac{\pi}{2^n}\right)$$

(exercice 6.12).

1. *En utilisant, pour $p \geq 1$, le développement limité à l'ordre $2p$ de la fonction* sin, *donner un développement asymptotique de la suite u.*
2. *En déduire les suites accélératrices correspondantes à la méthode de Richardson.*

Solution 6.19 *1. Avec le développement limité :*

$$\frac{\sin(\pi x)}{\pi x} = \sum_{j=0}^{p+1} (-1)^j \frac{\pi^{2j}}{(2j+1)!} x^{2j} + o\left(x^{2p+2}\right),$$

on obtient le développement asymptotique :

$$u_n = \pi + \sum_{j=1}^{p+1} (-1)^j \frac{\pi^{4j+1}}{(2j+1)!} \left(\frac{1}{4^j}\right)^n + o\left(\left(\frac{1}{4^{p+1}}\right)^n\right).$$

2. Les suites accélératrices correspondantes à la méthode de Richardson sont donc données par :

$$\begin{cases} u_{n,0} = u_n = 2^n \sin\left(\dfrac{\pi}{2^n}\right),\ (n \geq 1) \\ u_{n,k} = \dfrac{u_{n+1,k-1} - \frac{1}{4^k} u_{n,k-1}}{1 - \frac{1}{4^k}} = \dfrac{4^k u_{n+1,k-1} - u_{n,k-1}}{4^k - 1}\ (1 \leq k \leq p,\ n \geq 1). \end{cases}$$

L'exercice précédent peut se ramener à la situation suivante. On dispose d'une fonction f définie sur $I =]-1,1[$ et admettant un développement limité d'ordre $p+1$ en 0 :

$$f(x) = \ell + \sum_{j=1}^{p+1} \beta_j x^j + o\left(x^{p+1}\right)$$

où p est un entier naturel non nul et les coefficients β_j sont tous non nuls. On associe à cette fonction la suite $(u_n)_{n\geq 1}$ définie par :

$$\forall n \geq 1,\ u_n = f(r^n)$$

où r est un réel non nul dans $]-1,1[$. On a alors le développement asymptotique :

$$u_n = \ell + \sum_{j=1}^{p+1} \beta_j \lambda_j^n + o\left(\lambda_{p+1}^n\right)$$

où les coefficients $\lambda_j = r^j$ vérifient bien l'hypothèse :

$$0 < |\lambda_{p+1}| < |\lambda_p| < \cdots < |\lambda_1| < 1.$$

On peut donc définir les suites accélératrices $(u_{n,k})_{n\geq 1}$ par :

$$\begin{cases} u_{n,0} = u_n = f(r^n) \\ u_{n,k} = \dfrac{u_{n+1,k-1} - r^k u_{n,k-1}}{1 - r^k} \ (1 \leq k \leq p). \end{cases}$$

Les coefficients $\beta_{k,k+1}$ sont alors donnés par :

$$\begin{aligned}
\beta_{k,k+1} &= \beta_{k+1} \prod_{j=1}^{k} \frac{r^{k+1} - r^j}{1 - r^j} = \beta_{k+1} \prod_{j=1}^{k} r^j \frac{r^{k+1-j} - 1}{1 - r^j} \\
&= \frac{\left(r^k - 1\right)\left(r^{k-1} - 1\right)\cdots(r-1)}{(1-r)\cdots\left(1 - r^{k-1}\right)\left(1 - r^k\right)} \beta_{k+1} \prod_{j=1}^{k} r^j \\
&= (-1)^k r^{(1+2+\cdots+k)} \beta_{k+1} \ = \ (-1)^k r^{\frac{k(k+1)}{2}} \beta_{k+1}
\end{aligned}$$

On a donc, pour tout k compris entre 0 et p :

$$u_{n,k} - \ell \underset{+\infty}{\backsim} (-1)^k \beta_{k+1} r^{\frac{(k+1)(2n+k)}{2}}$$

Pour la suite d'Archimède, on a :

$$\forall n \geq 1,\ u_n = 2^n \sin\left(\frac{\pi}{2^n}\right) = f\left(\frac{1}{2^n}\right)$$

où $f(x) = \dfrac{\sin(\pi x)}{x}$ et en écrivant $f(x) = g\left(x^2\right)$, où :

$$g(t) = \pi + \sum_{j=1}^{p+1} (-1)^j \frac{\pi^{2j+1}}{(2j+1)!} t^j + o\left(x^{p+1}\right)$$

on se ramène à la situation précédente avec $r = \dfrac{1}{4}$. On a alors, pour tout k compris entre 0 et p :

$$\pi - u_{n,k} \underset{+\infty}{\backsim} \frac{\pi^{2k+3}}{(2k+3)!} \frac{1}{2^{(2n+k)(k+1)}}.$$

On a, par exemple, pour les trois premières suites :

$$\begin{cases} u_n = 2^n \sin\left(\dfrac{\pi}{2^n}\right), \\ u_{n,1} = \dfrac{4u_{n+1} - u_n}{3}, \\ u_{n,2} = \dfrac{16u_{n+1,1} - u_{n,1}}{15}, \end{cases}$$

avec :

$$\begin{cases} \pi - u_n \underset{+\infty}{\backsim} \dfrac{\pi^3}{6} \dfrac{1}{4^n}, \\ \pi - u_{n,1} \underset{+\infty}{\backsim} \dfrac{\pi^5}{5!} \dfrac{4}{16^{n+1}}, \\ \pi - u_{n,2} \underset{+\infty}{\backsim} \dfrac{\pi^7}{7!} \dfrac{1}{64^{n+1}}. \end{cases}$$

Exemple 6.2 *Si on reprend l'exemple du nombre e approché par la suite $(u_n)_{n\geq 0}$ définie par :*

$$\forall n \geq 0,\ u_n = \left(1 + \frac{1}{2^n}\right)^{2^n},$$

On a $u_n = f\left(\dfrac{1}{2^n}\right)$, où f est la fonction définie sur $]-1,1[$ par :

$$f(x) = \begin{cases} (1+x)^{\frac{1}{x}} = e^{\frac{\ln(1+x)}{x}} \text{ si } 0 < |x| < 1, \\ 0 \text{ si } x = 0. \end{cases}$$

Cette fonction est indéfiniment dérivable sur $]-1,1[$ comme composée des fonctions :

$$g : x \mapsto \frac{\ln(1+x)}{x} = \sum_{k=1}^{+\infty} (-1)^{n-1} \frac{x^{n-1}}{n}$$

et $t \mapsto e^t$*, elle admet donc des développements limités à tous ordres en* 0*. Par exemple, à l'ordre* 3*, on a :*

$$e^{\frac{\ln(1+x)}{x}} = e - \frac{e}{2}x + \frac{11e}{24}x^2 - \frac{7e}{16}x^3 + o\left(x^3\right).$$

On peut donc utiliser le procédé d'accélération de Richardson avec $r = \frac{1}{2}$*, ce qui donne les suites accélératrices définies par :*

$$\begin{cases} u_{n,0} = u_n = \left(1 + \dfrac{1}{2^n}\right)^{2^n} \ (n \geq 0), \\ u_{n,k} = \dfrac{2^k x_{n+1,k-1} - x_{n,k-1}}{2^k - 1} \ (k \geq 1), \end{cases}$$

et on a :

$$u_{n,k} - e \underset{+\infty}{\backsim} (-1)^k \beta_{k+1} \frac{1}{2^{(2n+k)\frac{k+1}{2}}}.$$

Ce qui donne, par exemple, pour les trois premières suites :

$$\begin{cases} u_n = \left(1 + \dfrac{1}{2^n}\right)^{2^n}, \\ u_{n,1} = 2u_{n+1} - u_n, \\ u_{n,2} = \dfrac{4u_{n+1,1} - u_{n,1}}{3}, \end{cases}$$

avec :

$$\begin{cases} e - u_n \underset{+\infty}{\backsim} \dfrac{e}{2}\dfrac{1}{2^n}, \\ e - u_{n,1} \underset{+\infty}{\backsim} \dfrac{11e}{24}\dfrac{1}{2^{2n+1}}, \\ e - u_{n,2} \underset{+\infty}{\backsim} \dfrac{7e}{16}\dfrac{1}{2^{3[n+1]}}. \end{cases}$$

Bibliographie

[1] B. Balaguer. *La leçon d'analyse au Capes de Mathématiques.* Ellipses (1999).

[2] M. Rogalski. *Carrefours entre analyse algèbre et géométrie.* Ellipses (2001).

[3] J. E. Rombaldi. *Problèmes corrigés d'analyse numérique.* Masson (1996).

[4] J. E. Rombaldi — *Éléments d'analyse réelle.* EDP Sciences (2004).

[5] J. E. Rombaldi — *Interpolation et approximation.* Vuibert (2005).

Chapitre 7

Matrices toutes puissantes

Matrices toutes puissantes.

(Arnaud de Saint Julien[1])

Résumé : Dans la RMS d'octobre 2005, Gabriel Dospinescu pose la question Q535 : déterminer les matrices carrées A telles que pour tout $n \in \mathbb{N}^*$, il existe une matrice B à coefficients rationnels telle que $A = B^n$. Nous allons répondre à cette question mais aussi proposer quelques prolongements.
Dans tout l'exposé, $\mathbb{K}$ désigne un corps commutatif. Une matrice carrée A est dite **toute puissante sur** $\mathbb{K}$ (en abrégé TP$\mathbb{K}$), si pour tout $n \in \mathbb{N}^*$, il existe une matrice B à coefficients dans $\mathbb{K}$ telle que $A = B^n$. On remarque déjà qu'une matrice toute puissante sur $\mathbb{K}$ est nécessairement à coefficients dans $\mathbb{K}$.
L'objectif de cet article est de déterminer les matrices toutes puissantes dans le cas où $\mathbb{K}$ désigne $\mathbb{C}$, $\mathbb{R}$, $\mathbb{Q}$ ou un corps fini.

Avant de démarrer je tiens à remercier vivement Vincent Beck pour sa précieuse relecture.

Mots-clés : bijection, unipotent, nilpotent, image de l'exponentielle de matrices complexes et réelles, réduction de Dunford, réduction simultanée, racines carrés de matrices, extensions cyclotomiques, corps finis.

7.1 Le cas instructif de la taille 1

Commençons par examiner ce qui se passe en taille 1, c'est-à-dire dans le cas des nombres.

[1] Lycée La Merci à Montpellier, France, desainta@yahoo.fr.

Sur $\mathbb{C}$ la situation est extrêmement simple. 0 est bien sûr TP$\mathbb{C}$ puisque pour tout $n \in \mathbb{N}^*$, $0 = 0^n$. Si $a \in \mathbb{C}^*$, $a = re^{i\theta} = e^{\ln r + i\theta}$. a est donc une exponentielle, ce qui permet d'écrire

$$a = (e^{\frac{\ln r + i\theta}{n}})^n.$$

Tous les nombres complexes sont donc TP$\mathbb{C}$.

Remarque : En fait sur un corps algébriquement clos, tous les nombres sont tout puissants puisque pour tout n l'équation $x^n = a$ admet une solution.

Sur $\mathbb{R}$, c'est presque aussi simple. Un nombre TP$\mathbb{R}$ est en particulier un carré. Réciproquement, si a non nul est un carré, $a > 0$ et $a = e^{\ln a} = (e^{\frac{\ln a}{n}})^n$. Les nombres TP$\mathbb{R}$ sont donc les nombres réels qui sont des carrés.

Sur $\mathbb{Q}$, la situation est plus délicate. Si $a \in \mathbb{Q}$ et $a > 0$, on peut toujours écrire $a = (e^{\frac{\ln a}{n}})^n$, le problème c'est que $e^{\frac{\ln a}{n}}$ n'a aucune raison d'être rationnel. Le théorème suivant nous apporte la solution.

Théorème 7.1 *Les nombres tous puissants sur* $\mathbb{Q}$ *sont* 0 *et* 1*.*

Preuve : 0 est bien sûr tout puissant puisque pour tout $n \in \mathbb{N}^*$, $0 = 0^n$.
Un nombre tout puissant a est en particulier un carré, il est donc positif.
Soit $a > 0$ un nombre tout puissant que l'on écrit en le décomposant en produit de facteurs premiers :

$$a = \prod_{p \in \mathcal{P}} p^{v_p(a)}$$

où $\mathcal{P}$ est l'ensemble des nombres premiers, et où les $v_p(a)$ sont des entiers relatifs presque tous nuls. Si on a $a = b^n$, on voit en décomposant b sous la même forme, que $v_p(a) = nv_p(b)$ (par unicité de la décomposition en produit de facteurs premiers), donc $v_p(a)$ est divisible par n et ce pour tout n puisque a est tout puissant, donc $v_p(a) = 0$ pour tout p, donc $a = 1$. Réciproquement 1 est bien tout puissant sur $\mathbb{Q}$ puisque pour tout $n \in \mathbb{N}^*$, $1 = 1^n$. □

Sur un corps fini $\mathbb{F}_q$, c'est simple aussi. Si $a \in \mathbb{F}_q$ est tout puissant non nul, il existe en particulier b non nul tel que $a = b^{q-1} = 1$ puisque le groupe des inversibles du corps $\mathbb{F}_q$ est cyclique d'ordre $q - 1$. Les nombres tout puissants sur un corps fini sont donc 0 et 1.

Donnons enfin une conséquence facile mais néanmoins très utile :

Proposition 7.1 *Le déterminant d'une matrice* TP$\mathbb{K}$ *est aussi* TP$\mathbb{K}$*.*

Preuve : Si A est TP$\mathbb{K}$, pour tout $n \in \mathbb{N}^*$ il existe une matrice B à coefficients dans $\mathbb{K}$ telle que $A = B^n$. Par multiplicativité du déterminant, on a donc $\det A = (\det B)^n$ avec $\det A$ et $\det B$ à coeffcients dans $\mathbb{K}$, d'où le résultat. □

Cette étude des nombres tout puissants nous laisse envisager que dans le cas général l'exponentielle de matrice va jouer un rôle et que sur $\mathbb{Q}$ des propriétés arithmétiques seront mises en jeu.

7.2 Généralités

7.2.1 Dévissage du problème par théorème spectral caractéristique

Dans tout cet l'exposé, p est un entier naturel non nul et I_p est la matrice unité de $M_p(\mathbb{K})$.
Le théorème de décomposition des noyaux et un argument de réduction simultanée vont nous permettre de dévisser les matrices toutes puissantes en blocs plus simples tout puissants. Le petit lemme suivant est à cet égard fondamental :

Lemme 7.1 *Soit E un $\mathbb{K}$-espace vectoriel et u et v deux endomorphismes de E qui commutent. Pour tout polynôme $Q \in \mathbb{K}[X]$, v laisse stable l'espace vectoriel $\operatorname{Ker} Q(u)$. En particulier, v laisse stable les sous-espaces propres et les sous-espaces caractéristiques de u.*

Preuve : Par commutation, $Q(u) \circ v = v \circ Q(u)$. Soit $x \in \operatorname{Ker} Q(u)$, $Q(u)(v(x)) = v(Q(u)(x)) = v(0) = 0$; ce qui prouve la stabilité. Si λ est une valeur propre de u de mutiplicité α, le lemme appliqué à respectivement $Q(x) = X - \lambda$ et $Q(x) = (X - \lambda)^\alpha$ montre que v laisse stable les sous-espaces propres et les sous espaces-caractéristiques de u. □

Remarque : Puisque u commute avec lui-même, on obtient en particulier que u stabilise ses sous-espaces propres et ses sous-espaces caratéristiques.

Proposition 7.2 (Théorème spectral caractéristique et dévissage) *Soit $A \in M_p(\mathbb{K})$ une matrice de polynôme caractéristique χ_A scindé sur le corps $\mathbb{K}$. Si $\chi_A = (X - \lambda_1)^{\alpha_1} \dots (X - \lambda_k)^{\alpha_k}$ alors,*

1) A est semblable sur $\mathbb{K}$ à la matrice diagonale par blocs

$$\operatorname{diag}(\lambda_1 I_{p_1} + N_1, \dots, \lambda_k I_{p_k} + N_k)$$

avec pour $i \in \{1, \dots, k\}, p_i = \dim \operatorname{Ker}(A - \lambda_i I_p)^{\alpha_i}$ et $N_i \in M_{p_i}(\mathbb{K})$ nilpotente.

2) A est TP$\mathbb{K}$ *si et seulement si pour tout $i \in \{1, \ldots, k\}$, $\lambda_i I_{p_i} + N_i$ est* TP$\mathbb{K}$*.*

Preuve : 1) On note u l'endomorphisme associé à A dans la base canonique de $\mathbb{K}^p$. En appliquant le lemme de décomposition des noyaux à χ_u, on a

$$\mathbb{K}^p = \operatorname{Ker}(u - \lambda_1 Id)^{\alpha_1} \oplus \ldots \oplus \operatorname{Ker}(u - \lambda_k Id)^{\alpha_k}.$$

Les $C_i = \operatorname{Ker}(u - \lambda_i Id)^{\alpha_i}$ sont les sous-espaces caractéristiques de u, ils sont stables par u. On peut donc définir $u_{|C_i}$ l'endomorphisme induit par u sur C_i. Pour tout $x \in C_i$, $(u - \lambda_i Id)^{\alpha_i}(x) = 0$ donc $u_{|C_i} - \lambda_i Id$ est nilpotent. Dans une base de C_i, la matrice de $u_{|C_i}$ s'écrit donc $\lambda_i I_{p_i} + N_i$ avec N_i nilpotente. Si $\mathcal{B}$ est une base de $\mathbb{K}^p$ obtenue en recollant des bases des C_i, on a

$$[u]_{\mathcal{B}} = \operatorname{diag}(\lambda_1 I_{p_1} + N_1, \ldots, \lambda_k I_{p_k} + N_k).$$

2) Si A est de plus TP$\mathbb{K}$, pour tout $n \in \mathbb{N}^*$ il existe $B \in M_p(\mathbb{K})$ telle que $A = B^n$. On note v l'endomorphisme associé à B dans la base canonique de $\mathbb{K}^p$. Puisque A est un polynôme en B, v et u commutent, v laisse stable les sous-espaces caractéristiques de u.
La matrice de v dans $\mathcal{B}$ (la base adpatée à la décomposition en sous-espaces caractéristiques de u) s'écrit donc

$$[v]_{\mathcal{B}} = \operatorname{diag}(B_1, \ldots, B_k)$$

avec $B_i \in M_{p_i}(\mathbb{K})$.
On a donc diagonalisé par bloc u et v dans une même base. Voici la traduction matricielle. Si P désigne la matrice de passage de la base canonique de $\mathbb{K}^p$ à $\mathcal{B}$, on a

$$A = P \operatorname{diag}(\lambda_1 I_{p_1} + N_1, \ldots, \lambda_k I_{p_k} + N_k) P^{-1} \text{ et } B = P \operatorname{diag}(B_1, \ldots, B_k) P^{-1}.$$

Comme $A = B^n$, par produit des blocs, on tire que pour tout $i \in \{1, \ldots, k\}$ et pour tout $n \in \mathbb{N}^*$,

$$\lambda_i I_{p_i} + N_i = {B_i}^n.$$

Les matrices $\lambda_i I_{p_i} + N_i$ sont donc TP$\mathbb{K}$.
Réciproquement, si les matrices $\lambda_i I_{p_i} + N_i$ sont TP$\mathbb{K}$, on a pour tout $n \in \mathbb{N}^*$

$$A = P \operatorname{diag}({B_1}^n, \ldots, {B_k}^n) P^{-1} = (P \operatorname{diag}(B_1, \ldots, B_k) P^{-1})^n,$$

ce qui prouve que A est TP$\mathbb{K}$. $\square$

Dans le cas où le polynôme caractéristique est **scindé**, la décomposition en sous-espaces caractéristiques permet donc de "dévisser" le problème. Il suffit

de déterminer les matrices toutes puissantes de la forme $\lambda I_p + N$ avec N nilpotente.
La réduction du théorème 7.2 va nous fournir en plus la décomposition de **Dunford** que nous utiliserons pour démontrer le lemme 7.2.

Corollaire 7.1 (Décomposition de Dunford) *Soit $A \in M_p(\mathbb{K})$ tel que son polynôme caractéristique χ_A est scindé sur $\mathbb{K}$. Il existe un unique couple (D, N) de matrices de $M_p(\mathbb{K})$ avec D diagonalisable sur $\mathbb{K}$ et N nilpotente tel que $A = D + N$ et $DN = ND$. De plus, D et N sont des polynômes en A.*

Preuve : 1) Reprenons les hypothèses du théorème précédent. On note $D = P\,\mathrm{diag}(\lambda_1 I_{p_1}, ..., \lambda_k I_{p_k})P^{-1}$ et $N = P\,\mathrm{diag}(N_1, ..., N_k)P^{-1}$, on a alors

$$A = D + N,$$

avec D diagonalisable sur $\mathbb{K}$ et N nilpotente. De plus, D et N commutent puisque les blocs N_i commutent avec les blocs d'homothétie $\lambda_i I_{p_i}$.

2) Montrons que D et N sont des polynômes en A. Pour tout i, on note p_i la projection sur C_i parallèlement à $\oplus_{j\neq i} C_j$ et D_i sa matrice dans la base $\mathcal{B}$. Comme $D = \sum_i \lambda_i D_i$, il suffit de montrer que les matrices D_i sont des polynômes en A.
On pose $M_i = (X - \lambda_i)^{\alpha_i}$ et $P_i = \prod_{j\neq i} M_j$. Puisque les M_i sont premiers deux à deux, aucun facteur n'est commun à tous les P_i, ils sont donc premiers dans leur ensemble. D'après Bezout, il existe donc des polynômes U_i tels que $\sum_i U_i P_i = 1$. Le polynôme $Q_i = U_i P_i$ vérifie alors :

$$(*) \quad \begin{cases} Q_i = 1 \mod (X - \lambda_i)^{\alpha_i} \\ \forall j \neq i \quad Q_i = 0 \mod (X - \lambda_j)^{\alpha_j}. \end{cases}$$

$Q_i(u)$ s'annule donc sur $\oplus_{j\neq i} C_j$ et coïncide avec l'identité sur C_i, c'est donc la projection sur C_i parallèlement à $\oplus_{j\neq i} C_j$, ce qui donne matriciellement $D_i = Q_i(A)$, et donc D est bien un polynôme en A, donc $N = A - D$ aussi.

Remarque : le théorème des restes chinois assure directement l'existence d'un polynôme Q_i vérifiant $(*)$.

3) Unicité : Si $A = D' + N'$, alors $D - D' = N - N'$. Comme ces 4 matrices sont des polynômes en A, D commute avec D' et N commute avec N', donc $D - D'$ est diagonalisable (elles sont simultanément diagonalisables car elles commutent) et $N - N'$ est nilpotente comme somme de matrices nilpotentes qui commutent (cela sera détaillé dans la preuve de la proposition 7.3).
Comme la seule matrice à la fois nilpotente et diagonalisable est la matrice nulle, on en déduit que $D - D' = N - N' = 0$, ce qui prouve l'unicité. $\square$

7.2.2 L'exponentielle : une bijection entre nilpotents et unipotents

Une matrice A appartenant à $M_p(\mathbb{K})$ est dite **unipotente** s'il existe une matrice $N \in M_p(\mathbb{K})$ nilpotente telle que $A = I_p + N$.
On note respectivement $\mathcal{U}_p(\mathbb{K})$ et $\mathcal{N}_p(\mathbb{K})$ l'ensemble des matrices de $M_p(\mathbb{K})$ unipotentes et nilpotentes.
Afin de pouvoir définir l'exponentielle et le logarithme de matrices, on va supposer ici que le corps $\mathbb{K}$ est de caractéristique nulle (on peut sinon pour simplifier les choses supposer que $\mathbb{K}$ est un sous-corps de $\mathbb{C}$).
On restreint la définition de l'exponentielle aux matrices nilpotentes.
Soit $N \in \mathcal{N}_p(\mathbb{K})$, on pose $\exp(N) = \sum_{n=0}^{+\infty} \frac{N^n}{n!}$ (somme de support finie).
Soit $U \in \mathcal{U}_p(\mathbb{K})$, on pose $\log U = \sum_{n=1}^{+\infty} (-1)^{n-1} \frac{(I-I_p)^n}{n}$ (somme de support finie).
Le lecteur aura remarqué que si le corps était de caractéristique p, on ne sait pas a priori ce que signifie $\frac{N^p}{p!}$, et donc on ne sait pas définir l'exponentielle d'une matrice nilpotente.

Proposition 7.3 *L'application* ***exponentielle*** *réalise une bijection entre les ensembles $\mathcal{N}_p(\mathbb{K})$ et $\mathcal{U}_p(\mathbb{K})$. De plus, l'application réciproque est la fonction* ***logarithme****.*

Preuve : 1) Si $N \in \mathcal{N}_p(\mathbb{K})$,

$$\exp N = I_p + \underbrace{\sum_{n=1}^{p-1} \frac{N^n}{n!}}_{N'}$$

puisque l'indice de nilpotence est inférieur ou égal à p. La matrice $\exp(N)$ est donc bien à coefficients dans $\mathbb{K}$.
N' est une somme finie de matrices nilpotentes qui commutent 2 à 2, c'est donc une matrice nilpotente. En effet si N_1 et N_2 sont dans $\mathcal{N}_p(\mathbb{K})$ et commutent, par le binôme de Newton, on a $(N_1 + N_2)^{2p} = \sum_{k=0}^{2p} \binom{2p}{k} N_1^k N_2^{2p-k} = 0$ car pour $k \geq p$, $N_1^k = 0$ et pour $k < p$, $N_2^{2p-k} = 0$.

2) La fonction logarithme est bien définie sur $\mathcal{U}_p(\mathbb{K})$ car c'est une somme finie. En effet, si $A \in \mathcal{U}_p(\mathbb{K})$, $(A - I_p)^k = 0$ dès que $k \geq p$. Ainsi

$$\log A = \sum_{n=1}^{p-1} (-1)^{n-1} \frac{(A - I_p)^n}{n}$$

est nilpotente puisque c'est une somme finie de matrices nilpotentes qui commutent 2 à 2.

3) Reste à prouver la bijection. Pour tout $x \in \mathbb{R}$, $\exp(x) = \sum_{n=0}^{+\infty} \frac{x^n}{n!}$, pour tout réel y tel que $|y-1| < 1$, on a $\ln(y) = \sum_{n=1}^{+\infty} (-1)^{n-1} \frac{(y-1)^n}{n}$. Comme $\exp(\ln y) = y$, on a l'identité suivante entre séries formelles :

$$\sum_{n=0}^{+\infty} \frac{1}{n!} \left(\sum_{k=1}^{+\infty} (-1)^{k-1} \frac{(Y-1)^k}{k} \right)^n = Y.$$

Si on prend pour Y une matrice $A \in \mathcal{U}_p(\mathbb{K})$, cela donne $\exp(\log A) = A$. On raisonne de même pour obtenir $\log(\exp N) = N$ si $N \in \mathcal{N}_p(\mathbb{K})$. □

Corollaire 7.2 *Les matrices unipotentes de $M_p(\mathbb{K})$ sont des exponentielles et sont toutes puissantes sur $\mathbb{K}$.*

Preuve : Si $A \in M_p(\mathbb{K})$ est unipotente, d'après la proposition précédente, il existe $N' \in M_p(\mathbb{K})$ nilpotente telle que $A = \exp(N')$.
Pour tout $n \in \mathbb{N}^*$, $A = [\exp(N'/n)]^n$ (car N'/n commute avec elle-même), avec $B = \exp(N'/n)$ qui est bien à coefficients dans $\mathbb{K}$ puisque N'/n est une matrice nilpotente à coefficients dans $\mathbb{K}$. □

Proposition 7.4 *La seule matrice nilpotente toute puissante sur $\mathbb{K}$ est la matrice nulle.*

Preuve : Si $N \in M_p(\mathbb{K})$ est TP$\mathbb{K}$, en particulier il existe une matrice R telle $N = R^p$. R est alors nilpotente et donc $R^p = 0$ ce qui donne $N = 0$, qui est bien sûr toute puissante puisque pour tout $n \in \mathbb{N}^*$, $0 = 0^n$. □

7.2.3 Le cas non inversible se ramène au cas inversible

Dans ce paragraphe, on montre que pour déterminer les matrices TP$\mathbb{K}$, il suffit de s'intéresser aux matrices inversibles. La proposition suivante précise le propos.

Proposition 7.5 *Une matrice A non inversible est* TP$\mathbb{K}$ *si et seulement si il existe une matrice M inversible et* TP$\mathbb{K}$ *telle que A soit semblable sur $\mathbb{K}$ à la matrice par blocs* $\text{diag}(0, M)$.

Preuve : 1) Soit A non inversible TP$\mathbb{K}$. On note u et v les endomorphismes respectivement associés à A et B dans la base canonique de $\mathbb{K}^p$. Le noyau de u n'est pas réduit à 0. Ainsi $\chi_u = X^r Q(X)$ où X^r et Q sont premiers entre eux

et r est non nul.
Le lemme de décomposition des noyaux donne $\mathbb{K}^p = \operatorname{Ker} u^r \oplus \operatorname{Ker} Q(u)$.
Puisque u et v commutent, v laisse stable $\operatorname{Ker} u^r$ et $\operatorname{Ker} Q(u)$. En choisissant une base $\mathcal{B}$ de $\mathbb{K}^p$ adaptée à la somme directe ci-dessus, et en notant P la matrice de passage de la base canonique à $\mathcal{B}$, on a comme dans la preuve de 7.2

$$A = P \operatorname{diag}(A_1, A_2) P^{-1} \quad \text{et} \quad B = P \operatorname{diag}(B_1, B_2) P^{-1},$$

avec en plus A_1 nilpotente et A_2 inversible.
Pour tout $n \in \mathbb{N}^*$ $A_i = B_i^n$ donc les matrices A_i sont TP$\mathbb{K}$.
A_1 est nilpotente et TP$\mathbb{K}$, elle est donc nulle d'après 7.4.
On conclut en posant $M = A_2$.

2) Réciproquement, supposons que $A = P \operatorname{diag}(0, M) P^{-1}$ avec M inversible et TP$\mathbb{K}$. Déjà A est non inversible. Si pour tout $n \in \mathbb{N}^*$ il existe B telle que $M = B^n$, on a

$$A = P \operatorname{diag}(0, B^n) P^{-1} = (P \operatorname{diag}(0, B) P^{-1})^n,$$

ce qui prouve que A est bien TP$\mathbb{K}$. □

7.3 Matrices toutes puissantes sur $\mathbb{C}$

D'après la proposition 7.5, il suffit de déterminer toutes les matrices inversibles toutes puissantes. Nous allons voir que, comme dans le cas de la dimension 1, la situation est assez simple : nous allons démontrer que toute matrice de $GL_p(\mathbb{C})$ est une exponentielle et est donc TP$\mathbb{C}$.

Proposition 7.6 *L'application exponentielle est une surjection de $M_p(\mathbb{C})$ sur $GL_p(\mathbb{C})$.*

Preuve : - Montrons déjà qu'une exponentielle est inversible. Toute matrice $M \in M_p(\mathbb{C})$ est triangularisable donc semblable à une matrice triangulaire dont la diagonale est constituée des valeurs propres de M. La matrice $\exp M$ est donc semblable à une matrice triangulaire dont la diagonale est constituée des exponentielles des valeurs propres de M. En passant au déterminant, on a donc

$$\det(\exp M) = \exp(\operatorname{Tr} M),$$

ce qui prouve que $\exp(M)$ est inversible.

- Si $A \in GL_p(\mathbb{C})$, 0 n'est pas valeur propre et d'après le théorème 7.2 (χ_A est scindé sur $\mathbb{C}$),

$$A = P \operatorname{diag}(\lambda_1 I_{\alpha_1} + N_1, \ldots, \lambda_k I_{\alpha_k} + N_k) P^{-1}.$$

Comme $\lambda_i \neq 0$, il existe $r_i \in \mathbb{C}$ tel que $\lambda_i = e^{r_i}$ ($\exp : \mathbb{C} \to \mathbb{C}^*$ est surjective). De plus il existe une matrice N_i' telle que $I_{\alpha_i} + \frac{N_i}{\lambda_i} = \exp N_i'$ puisqu'elle est unipotente. On a alors

$$\lambda_i I_{\alpha_i} + N_i = \lambda_i (I_{\alpha_i} + \frac{N_i}{\lambda_i}) = e^{r_i} \exp(N_i') = \exp(r_i I_{\alpha_i} + N_i')$$

car $r_i I_{\alpha_i}$ et N_i' commutent. Ainsi,

$$A = \exp\left(P \operatorname{diag}(r_1 I_{\alpha_1} + N_1', \ldots, r_k I_{\alpha_k} + N_k') P^{-1}\right). \quad \square$$

Remarque : L'exponentielle n'est pas bijective puisque ce n'est pas vrai en dimension 1.

La proposition précédente nous permet de déterminer toutes les matrices TP$\mathbb{C}$.

Théorème 7.2 (Description des matrices TP$\mathbb{C}$**)** *Les matrices toutes puissantes sur* $\mathbb{C}$ *sont les matrices* A *telles que* $\dim \operatorname{Ker} A = \alpha_0$ *où* α_0 *est la multiplicité de 0 dans le polynôme caractéristique de* A*. En particulier, toute matrice inversible est toute puissante sur* $\mathbb{C}$*.*

Preuve : - La proposition précédente implique directement que toute matrice inversible est TP$\mathbb{C}$. Dans ce cas $\dim \operatorname{Ker} u = 0$, qui est la multiplicité de la valeur propre 0 puisque 0 n'est pas valeur propre.

- Si A est une matrice non inversible et TP$\mathbb{C}$, d'après la proposition 7.5 et la proposition précédente, A est semblable à $\operatorname{diag}(0, M)$ avec M inversible. On note s la taille du bloc nul. Il est clair que la multiplicité de la valeur propre 0 dans χ_A est s et que $\dim \operatorname{Ker} A = s$. $\square$

Exemples :

1. La matrice $A = \begin{pmatrix} 0 & i & 7 \\ 0 & 0 & 3 \\ 0 & 0 & 5 \end{pmatrix}$ n'est pas TP$\mathbb{C}$ puisque 0 est valeur propre de multiplicité 2 mais que $\operatorname{rang} A = 2$ et donc $\dim \operatorname{Ker} A = 1 \neq 2$.

2. La matrice

$$M = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & -3 \end{pmatrix}$$

est la **matrice compagnon** associée au polynôme

$$P(X) = X^4 + 3X^3 + 3X^2 + X = X(X+1)^3.$$

On a donc $\chi_M = X(X+1)^3$. 0 est donc valeur propre de multiplicité 1. Comme rang$M = 3$, $\dim \operatorname{Ker} M = 1$ qui est la multiplicité de 0. M est donc TP$\mathbb{C}$.
M est à coefficients réels, on peut donc naturellement se demander si elle est TP$\mathbb{R}$. Comme χ_M est scindé sur $\mathbb{R}$, la décomposition en sous-espaces caractéristiques donne que M est semblable à $\operatorname{diag}(0, -I_3 + N)$ avec $N \in M_3(\mathbb{R})$ nilpotente. D'après la propriété 7.2 de dévissage, si M est TP$\mathbb{R}$, alors $-I_3 + N$ est aussi TP$\mathbb{R}$. Mais $\det(-I_3 + N) = (-1)^3 = -1 < 0$, donc $-I_3 + N$ n'est pas un carré et donc n'est pas TP$\mathbb{R}$. La matrice M est donc TP$\mathbb{C}$ mais n'est pas TP$\mathbb{R}$.

7.4 Matrices toutes puissantes sur $\mathbb{R}$

7.4.1 Premières constatations

Puisque toute exponentielle de matrice réelle est TP$\mathbb{R}$, il est naturel de déterminer l'image $\exp(M_p(\mathbb{R}))$. On a vu à la section 7.3 que pour tout $M \in M_p(\mathbb{C})$,

$$\det(\exp M) = \exp(\operatorname{Tr} M).$$

Cette égalité est donc en particulier vraie si $M \in M_p(\mathbb{R})$, ce qui donne $\det(\exp M) > 0$. Ainsi $\exp(M_p(\mathbb{R})) \subset GL_p^+(\mathbb{R})$. Cette inclusion n'est pas une égalité. En effet,

$$A = \begin{pmatrix} -3 & 0 \\ 0 & -4 \end{pmatrix}$$

n'est pas un carré, donc pas une exponentielle, pourtant $\det M > 0$. En effet, si $R^2 = A$, alors R commute avec A diagonale à spectre simple, donc R est aussi diagonale et s'écrit $R = \operatorname{diag}(r_1, r_2)$ avec $r_1^2 = -3$, ce qui est impossible.
Attention, il existe des matrices qui sont des exponentielles qui ont pourtant leurs valeurs propres négatives. Penser à

$$M = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},$$

la matrice de rotation d'angle π. Elle est le carré de la matrice de rotation d'angle $\pi/2$, $M = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^2$.
Plus généralement, il est facile de voir que M est TP$\mathbb{R}$. En effet, pour tout $n \in \mathbb{N}^*$,

$$M = \begin{pmatrix} \cos \pi/n & -\sin \pi/n \\ \sin \pi/n & \cos \pi/n \end{pmatrix}^n.$$

Ce résultat est naturel, si on itère n fois une rotation d'angle π/n, on obtient une rotation d'angle π. Par ailleurs, on peut montrer que pour tout réel θ,

$$\exp\begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

ce qui montre que M est aussi une exponentielle, $M = \exp\begin{pmatrix} 0 & -\pi \\ \pi & 0 \end{pmatrix}$.

7.4.2 Image de l'exponentielle de matrices réelles

En dimension 1, un nombre réel est une exponentielle si, et seulement si, c'est un carré non nul. Nous allons voir que cela se généralise en dimension supérieure, c'est l'objet du théorème suivant.

Théorème 7.3 (Image de l'exponentielle de matrice réelle)

$$\exp(M_p(\mathbb{R})) = \left\{R^2,\ R \in GL_p(\mathbb{R})\right\}.$$

Autrement dit, une matrice réelle inversible est une exponentielle si, et seulement si, c'est un carré. Ce résultat est assez remarquable : pour savoir si une matrice inversible réelle est une exponentielle, il suffit de tester si c'est un carré, ce qui est généralement plus simple.
La preuve que nous allons donner est tirée d'un article de Michel Coste [www], enseignant à la prépa-agreg de Rennes. Nous aurons auparavant besoin de démontrer un lemme qui raffine la surjectivité de $\exp : M_p(\mathbb{C}) \to GL_p(\mathbb{C})$.

Lemme 7.2 ***(Raffinement de la surjectivité de l'exponentielle complexe)*** *Pour toute matrice M de $GL_p(\mathbb{C})$, il existe un polynôme $P \in \mathbb{C}[X]$ tel que $M = \exp(P(M))$.*

Preuve : - Soit $M \in GL_p(\mathbb{C})$. On écrit sa décomposition de Dunford (χ_M est scindé sur $\mathbb{C}$) : $M = D + N$ avec D diagonalisable sur $\mathbb{C}$, N nilpotente, et D et N qui commutent. On utilisera en plus que D et N sont des polynômes en M, c'est un ingrédient essentiel.
Puisque D est inversible, $M = D(I_p + D^{-1}N)$. L'idée est de trouver deux polynômes U et V tels que $D = \exp(U(M))$ et $I_p + D^{-1}N = \exp(V(M))$. Puisque $U(M)$ et $V(M)$ commutent, on aura alors $M = \exp(U(M) + V(M))$ ce qui démontrera le lemme avec $P = U + V$.

- On écrit $D = P\,\mathrm{diag}(\lambda_1, \ldots, \lambda_p)P^{-1}$ où les λ_i sont les valeurs propres toutes non nulles puisque M est inversible.
Pour tout $\lambda \in \mathrm{Spec}(M)$, il existe un nombre complexe μ_λ tel que $\exp\mu_\lambda = \lambda$ puisque $\exp : \mathbb{C} \to \mathbb{C}^*$ est surjective. Il existe alors un polynôme interpolateur

de Lagrange L tel que pour tout $\mu \in \mathrm{Spec}(M)$, $L(\lambda) = \mu_\lambda$ puisque les λ sont disctincts deux à deux. Alors

$$\begin{aligned} D &= P\,\mathrm{diag}(\lambda_1, \dots, \lambda_p)P^{-1} \\ &= P\,\mathrm{diag}(\exp \mu_1, \dots, \exp \mu_p)P^{-1} \\ &= \exp\left(P\,\mathrm{diag}(\mu_1, \dots, \mu_p)P^{-1}\right) \\ &= \exp\left(P\,\mathrm{diag}(L(\lambda_1), \dots, L(\lambda_p))P^{-1}\right) \\ &= \exp\left((L(P\,\mathrm{diag}(\lambda_1, \dots, \lambda_p)P^{-1}\right) \\ &= \exp(L(D)). \end{aligned}$$

Comme D est un polynôme en M, $L(D)$ est aussi un polynôme en M, ce qui donne l'existence du polynôme U.

- Montrons que $I_p + D^{-1}N$ est unipotente. La matrice D^{-1} est un polynôme en D. En effet, d'après le théorème de Cayley-Hamilton, $\chi_D(D) = 0$, ce qui donne

$$(-1)^p D^p + a_{p-1}D^{p-1} + ... + a_1 D + a_0 I_p = 0$$

avec $a_0 = \det D \neq 0$. En composant par D^{-1} puis en divisant par a_0, on a donc

$$D^{-1} = a_0{}^{-1}\left((-1)^p D^{p-1} + a_{p-1}D^{p-2} + \dots + a_1 I_p\right),$$

ce qui donne le résultat. La matrice N commute avec D donc avec D^{-1} puisque c'est un polynôme en D.
On a alors $(D^{-1}N)^p = (D^{-1})^p N^p = 0$ ce qui prouve que $D^{-1}N$ est nilpotente et donc que $I_p + D^{-1}N$ est unipotente. D'après la proposition 7.3, elle est l'exponentielle d'un logarithme, plus précisément,

$$I_p + D^{-1}N = \exp\left(\sum_{k=1}^{p} \frac{(-1)^{k-1}}{k}(D^{-1}N)^k\right).$$

Ce logarithme est une somme finie, c'est un polynôme en $D^{-1}N$, donc un polynôme en M car D^{-1} et N le sont aussi. $I_p + D^{-1}N$ est donc un polynôme en M, ce qui donne l'existence du polynôme V et achève la preuve du lemme.
□

Preuve du Théorème 7.3 :

- Si $M = \exp(T)$, avec T réelle, M est inversible et $M = (\exp(T/2))^2$ donc est un carré.

- Passons à la réciproque. On suppose que $M = R^2$ avec $R \in GL_p(\mathbb{R})$. On applique le lemme 7.2 à R, il existe un polynôme complexe P tel que $R = \exp(P(R))$. Puisque R est réelle, en passant au conjugué, on a aussi

$R = \exp(\overline{P}(R))$. $P(R)$ et $\overline{P}(R)$ commutent, donc $R^2 = \exp(P(R) + \overline{P}(R))$, ce qui montre que $M = R^2$ est l'exponentielle de la matrice réelle $P(R) + \overline{P}(R)$. $\square$

Remarque : Contrairement à la dimension 1, exp **n'est pas injective** sur $M_p(\mathbb{R})$ pour $p \neq 1$. Par exemple pour $p = 2$, trouvons une matrice réelle qui a pour valeurs propres $2i\pi$ et $-2i\pi$. Le polynôme caractéristique vaut alors $X^2 + 4\pi^2$, il n'y a qu'à prendre sa matrice compagnon

$$A = \begin{pmatrix} 0 & -4\pi^2 \\ 1 & 0 \end{pmatrix}.$$

A est diagonalisable sur $\mathbb{C}$, semblable à $\mathrm{diag}(2i\pi, -2i\pi)$ donc son exponentielle est semblable à $\mathrm{diag}(\exp(2i\pi), \exp(-2i\pi)) = I_2$. On a donc

$$\exp(A) = \exp(0) = I_2,$$

ce qui montre qu'il n'y a pas injectivité.

7.4.3 Caractérisation des matrices TP$\mathbb{R}$

Compte tenu de la proposition 7.5, nous avons donc directement le résultat suivant :

Théorème 7.4 (Caractérisation des matrices TP$\mathbb{R}$)
Les matrices TP$\mathbb{R}$ *sont :*
- *les matrices de la forme B^2 avec B réelle inversible,*
- *les matrices semblables sur $\mathbb{R}$ à $\mathrm{diag}(0, M^2)$ avec M réelle inversible.*

Exemples : détermination des matrices TP$\mathbb{R}$ de $M_2(\mathbb{R})$

Soit $A \in M_2(\mathbb{R})$.

1. Supposons que χ_A est scindé sur $\mathbb{R}$.

a. Si les valeurs propres sont distinctes, alors A est diagonalisable sur $\mathbb{R}$, et admet une racine carrée si, et seulement si, ses deux valeurs propres sont positives ou nulles (l'idée est la même que pour $A = \begin{pmatrix} -3 & 0 \\ 0 & -4 \end{pmatrix}$, on peut quand même voir par exemple le sujet d'algèbre MP des CCP de 2005).

b. Si la valeur propre est double, A est semblable à $\lambda I_2 + N$ avec N nilpotente.

- Si $\lambda = 0$, A est nilpotente. Elle est donc TP$\mathbb{R}$ si, et seulement si, $A = 0$.

- Si $\lambda > 0$, A/λ est unipotente donc TP$\mathbb{R}$ et donc un carré R^2. Mais alors $A = (\sqrt{\lambda}R)^2$, donc A est TP$\mathbb{R}$.

- Si $\lambda < 0$, $\lambda = -\lambda'$ avec $\lambda' > 0$. Alors A/λ' est semblable soit à $-I_2$, soit à $A' = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$.

On a déja vu que $-I_2$ est un carré (le carré de la rotation d'angle $\pi/2$). Supposons que $A' = R^2$ avec R réelle, alors en trigonalisant simultanément A' et R, on voit que les valeurs propres de A' sont les carrés de celles de R, donc celles-ci valent i ou $-i$. Comme R est réelle, ses valeurs propres sont conjuguées donc R admet 2 valeurs propres distinctes, donc est diagonalisable sur $\mathbb{C}$, donc A' aussi, ce qui n'est pas. A' n'est donc pas un carré.

2. Supposons que χ_A ne soit pas scindé sur $\mathbb{R}$. Alors c'est un polynôme irréductible sur $\mathbb{R}$, donc A est semblable sur $\mathbb{R}$ à une matrice du type

$$S = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

avec $b \neq 0$. Si on pose $z = a + ib = re^{i\theta}$, S représente dans la base canonique la matrice de la similitude directe de rapport r et d'angle θ :

$$S = rR_\theta \quad \text{avec } R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

S est donc le carré par exemple de la matrice $\sqrt{r}R_{\theta/2}$.

7.4.4 Caractérisation des carrés inversibles

Le théorème 7.4 ne peut être satisfaisant, que si l'on sait reconnaître les matrices réelles inversibles qui sont des carrés. Le théorème suivant énoncé par Rached Mneimné dans [Mn] p. 90 apporte une réponse relativement simple à notre problème.

Théorème 7.1 *Une matrice réelle inversible est un carré si, et seulement si, pour chaque (éventuelle) valeur propre négative, les blocs de Jordan associés de même taille sont en nombre pair.*

En particulier, nous obtenons :

Théorème 7.2 *Une matrice réelle qui n'a pas de valeur propre négative est* TP$\mathbb{R}$.

Exemples : 1. La matrice

$$A = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

est TP$\mathbb{R}$ puisque $\chi_A = X^4 + 2X^2 + 1 = (X^2 + 1)^2$.

2. Déterminons les matrices TP$\mathbb{R}$ de la forme $A = -I_4 + N$ avec $N \in M_4(\mathbb{R})$ nilpotente. On considère les blocs de Jordan

$$J_k(\lambda) = \begin{pmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix} \in M_k(\mathbb{R}) \quad \text{avec } J_1(\lambda) = (\lambda).$$

Les résultats sur la réduction de Jordan disent que A est semblable à l'une des matrices suivantes : $A_1 = \text{diag}(-1, -1, -1, -1)$, $A_2 = \text{diag}(-1, -1, J_2(-1))$, $A_3 = \text{diag}(-1, J_3(-1))$, $A_4 = \text{diag}(J_2(-1))$, $J_2(-1))$ ou $A_5 = J_4(-1)$.
Parmi ces matrices, les seules qui sont TP$\mathbb{R}$ sont $A_1 = \text{diag}(-1, -1, -1, -1)$ et $A_4 = \text{diag}(J_2(-1))$, $J_2(-1))$.

7.5 Matrices toutes puissantes sur un corps fini

Dans cette section $\mathbb{K}$ désigne le corps $\mathbb{F}_q$. Remarquons que le groupe $GL_p(\mathbb{F}_q)$ est fini, notons N son cardinal. Soit M une matrice inversible toute puissante sur $\mathbb{F}_q$, il existe une matrice B de $GL_p(\mathbb{F}_q)$ telle que $M = B^N$. Mais par le théorème de Lagrange, $B^N = I_p$ ce qui donne $M = I_p$. Grâce à la proposition 7.5, on vient donc de prouver qu'une matrice toute puissante est semblable à une matrice diagonale avec uniquement 0 et 1 sur la diagonale, c'est-à-dire la matrice d'un projecteur. Réciproquement, une telle matrice M est bien toute puissante car pour tout $n \in \mathbb{N}^*$, $M = M^n$. On a donc prouvé :

Théorème 7.5 *Les matrices toutes puissantes sur un corps fini sont les matrices des projecteurs.*

7.6 Matrices toutes puissantes sur $\mathbb{Q}$

7.6.1 Cas des matrices de la forme $rI_p + N$

Nous allons chercher quelles sont les matrices TP$\mathbb{Q}$ de la forme $rI_p + N$ avec $r \in \mathbb{Q}$ et $N \in \mathcal{N}_p(\mathbb{Q})$. Supposons que la matrice $A = rI_p + N$ soit TP$\mathbb{Q}$. Alors son déterminant qui vaut r^p est égal à 0 ou 1 car TP$\mathbb{Q}$.

• Si $r^p = 0$ alors $r = 0$ et $A = N$ est nilpotente et TP$\mathbb{Q}$ donc $A = 0$ d'après la proposition 7.4.

• Si $r^p = 1$ comme $r \in \mathbb{Q}$, on a $r = 1$ ou -1 et dans ce cas p est pair.

- La condition $r = 1$ signifie que A est unipotente. Réciproquement, d'après le corollaire 7.2 une matrice unipotente est bien TP$\mathbb{Q}$.

- Reste à traiter le cas $r = -1$. Nous allons démontrer qu'une matrice du type $-I_p + N$ n'est pas TP$\mathbb{Q}$. Pour cela nous aurons besoin du lemme suivant.

Lemme 7.3 *Pour tout entier naturel k, le polynôme* $X^{2^k} + 1$ *est irréductible sur* $\mathbb{Q}$*.*

Preuve : • **Première démonstration :** On pose $P(X) = X^{2^k} + 1$. L'idée est d'appliquer le critère d'Eisenstein ([Pe] p 76) au polynôme $P(X + 1)$.

- Montrons par récurrence sur k que tous les coefficients de $(X + 1)^{2^k}$ mis à part le premier et le dernier sont pairs. C'est vrai pour $k = 1$ puisque $(X + 1)^2 = X^2 + 2X + 1$. Supposons que c'est vrai au rang k.

$$(X+1)^{2^{k+1}} = ((X+1)^{2^k})^2 = \left(\sum_{i=0}^{2^k} b_i X^i\right)^2 = \sum_{i=0}^{2^k} b_i^2 X^{2i} + 2 \sum_{\substack{i,j \in \{0,\ldots,2^k\} \\ i<j}} b_i b_j X^{i+j}.$$

Par hypothèse de récurrence, pour $i \in \{1, ..., 2^k - 1\}$, b_i est pair. Si l'on regarde modulo 2, on a donc

$$(X + 1)^{2^{k+1}} \equiv \sum_{i=0}^{2^k} b_i^2 X^{2i} \equiv X^{2^{k+1}} + 1$$

ce qui prouve l'hérédité.

- Le polynôme $P(X + 1) = (X + 1)^{2^k} + 1$ a donc comme terme de plus haut degré X^{2^k}, son coefficient constant vaut 2 et tous les autres coefficients sont pairs. Il vérifie donc le critère d'Eisenstein avec $p = 2$ et est donc irréductible sur $\mathbb{Q}$.

- Si $P(X)$ est réductible sur $\mathbb{Q}$, $P(X) = U(X)V(X)$ avec U et V polynômes à coefficients dans $\mathbb{Q}$ de degrés non nuls. Mais alors $P(X+1) = U(X+1)V(X+1)$ avec $U(X + 1)$ et $V(X + 1)$ de degrés non nuls et à coefficients dans $\mathbb{Q}$, ce qui est impossible puisque $P(X + 1)$ est irréductible sur $\mathbb{Q}$. $P(X) = X^{2^k} + 1$ est donc bien irréductible sur $\mathbb{Q}$. □

• **Deuxième démonstration :** Montrons que $X^{2^k} + 1$ est en fait $\phi_{2^{k+1}}$ le polynôme cyclotomique d'ordre 2^{k+1}, comme les polynômes cyclotomiques sont irréductibles sur $\mathbb{Q}$ ([Pe] p 82), cela démontrera le lemme.

On rappelle que $\phi_{2^{k+1}} = \prod(X - \zeta)$ où ζ parcourt l'ensemble des racines primitives 2^{k+1}-ièmes de l'unité. Le polynôme $\phi_{2^{k+1}}$ est donc unitaire, et son degré vaut $\varphi(2^{k+1}) = 2^{k+1} - 2^k = 2^k$ où φ est la fonction indicatrice d'Euler. Soit ζ une racine primitive 2^{k+1}-ièmes de l'unité. On a

$$(\zeta^{2^k})^2 = \zeta^{2^{k+1}} = 1,$$

donc ζ^{2^k} est une racine primitive 2-ième de l'unité, donc $\zeta^{2^k} = -1$, ce qui entraîne que ζ est racine de $X^{2^k} + 1$. $\phi_{2^{k+1}}$ divise donc $X^{2^k} + 1$ mais comme ils sont unitaires et de même degré, ils sont égaux. □

Proposition 7.7 *Il n'existe aucune matrice* TP$\mathbb{Q}$ *de la forme* $-I_p + N$ *avec* $N \in \mathcal{N}_p(\mathbb{Q})$.

Preuve : Soit k un entier tel que $2^k > p$. Supposons que $A = -I_p + N$ est TP$\mathbb{Q}$, il existe en particulier $R \in M_p(\mathbb{Q})$ (une racine 2^k-ième) telle que $A = R^{2^k}$. Puisque N est nilpotente, on a $(A + I_p)^p = 0$ et $(R^{2^k} + I_p)^p = 0$ Le polynôme minimal de R noté π_R divise donc $(X^{2^k} + 1)^p$. Mais $X^{2^k} + 1$ est irréductible sur $\mathbb{Q}$ donc π_R est une puissance de $X^{2^k} + 1$, mais comme $\deg \pi_R \leq p$, nécessairement $\pi_R = 1$ ce qui est impossible puisqu'un polynôme minimal est toujours de degré supérieur ou égal à 1. □

Remarque : La preuve de cette proposition est instructive car la même tactique nous permettra de prouver le théorème plus général 7.8.

Compte tenu des résultats obtenus à la section 7.2.1, nous avons démontré le théorème suivant :

Théorème 7.6 *Soit A une matrice* TP$\mathbb{Q}$ *telle que son polynôme caractéristique est scindé sur* $\mathbb{Q}$.

a) Si A est inversible, A est unipotente.

b) Sinon A est semblable sur $\mathbb{Q}$ à la matrice par bloc $\mathrm{diag}(I_r + N, 0)$ *où N est une matrice rationnelle nilpotente et r le rang de la matrice.*

7.6.2 Une première étude du spectre d'une matrice TP$\mathbb{Q}$

L'objectif de cette partie est de montrer que les valeurs propres complexes d'une matrice TP$\mathbb{Q}$ sont nécessairement rationnelles, ce qui prouvera que son polynôme caractéristique est scindé sur $\mathbb{Q}$.
Une importante avancée nous est permise grâce au théorème suivant dont on trouvera une preuve de Daniel Perrin (je l'en remercie vivement) en annexe dans la section 7.6.4.

Théorème 7.7 (Perrin) *Soit $\mathbb{K}$ une extension de $\mathbb{Q}$ de degré finie. Les seuls nombres* TP$\mathbb{K}$ *sont 0 et 1.*

La preuve utilise de la théorie algébrique des nombres, elle utilise notamment deux résultats difficiles :

- les idéaux fractionnaires d'un anneau des entiers admettent une décomposition unique en produit d'idéaux premiers.

- le théorème des unités de Dirichlet qui précise la structure du groupe abélien des unités d'un corps de nombres.

Cette preuve est toutefois très instructive et culturelle, elle permet d'illustrer le fait qu'un idéal généralise la notion de nombre et qu'on peut y définir des opérations...

Conséquences : Le polynôme caractéristique χ_A est scindé sur son corps de décomposition noté $\mathbb{K}$. On peut donc diagonaliser A en blocs TP$\mathbb{K}$ de la forme $\lambda I_k + N$. Si λ est une valeur propre de A, et que k est la taille du bloc associé $\lambda I_k + N$, on obtient en passant au déterminant que λ^k est TP$\mathbb{K}$, donc vaut 0 ou 1 grâce au dernier théorème ; ce qui donne que λ vaut 0 ou bien est une racine de l'unité.
Si les valeurs propres sont simples, $k = 1$, et dans ce cas on conclut directement que λ vaut 0 ou 1. Nous venons donc d'obtenir :

Proposition 7.8 *Les valeurs propres d'une matrice* TP$\mathbb{Q}$ *sont parmi 0,1 ou les racines de l'unité.*

Nous allons maintenant prouver que parmi les racines de l'unité, 1 est la seule valeur propre possible pour une matrice TP$\mathbb{Q}$.

7.6.3 Etude finale du spectre et conclusion

A. Une famille de polynômes irréductibles

Nous allons prouver la proposition suivante (je remercie à ce sujet Clément de Seguin Pazis qui a permis de simplifier la preuve) :

Proposition 7.9 *Soit ζ une racine primitive n-ième de l'unité ($n \geq 2$). Pour tout entier $a \geq 1$, le polynôme $X^{n^a} - \zeta$ est irréductible sur $\mathbb{Q}(\zeta)$.*

Preuve : On rappelle que si ζ est une racine primitive n-ième de l'unité, $\mathbb{Q}(\zeta)$ est un $\mathbb{Q}$-espace vectoriel de dimension $\varphi(n)$ où φ est l'indicatrice d'Euler ([Pe] p. 83).

1. Soit z une racine de $P_a = X^{n^a} - \zeta$. En élevant à la puissance n, on voit que z est une racine n^{a+1}-ième de l'unité. Montrons qu'elle est en plus primitive. On sait déjà que l'ordre de z divise n^{a+1}.
Si r est un entier, on sait que l'ordre de z^r est égal à l'ordre de z divisé par pgcd(r,ordre(z)). Si l'on note k l'ordre de z, l'égalité $z^{n^a} = \zeta$ entraîne que les ordres de ces deux éléments sont égaux, et donc que

$$\frac{k}{\operatorname{pgcd}(n^a, k)} = n,$$

soit $k = n \operatorname{pgcd}(n^a, k)$. D'après Bezout, il existe des entiers u et v tels que $k = un^{a+1} + vnk$ $(*)$.
Supposons $k < n^{a+1}$, alors $n^{a+1}/k = q \geq 2$, et en divisant $(*)$ par k, on obtient $1 = uq + vn$, donc $\operatorname{pgcd}(q, n) = 1$, ce qui est impossible car $q \mid n^{a+1}$ et donc admet un diviseur premier qui divise n.
Les racines de P_a sont donc des racines primitives n^{a+1}-ièmes de 1, le corps de décomposition de P_a sur $\mathbb{Q}(\zeta)$ est donc le corps cyclotomique $\mathbb{Q}(z)$ avec z racine primitive n^{a+1}-ième de 1 dont la puissance n^a-ième vaut ζ.

2. L'extension cyclotomique $\mathbb{Q}(z)$ est de degré $\varphi(n^{a+1})$ au dessus de $\mathbb{Q}$, ie $[\mathbb{Q}(z) : \mathbb{Q}] = \varphi(n^{a+1}) = n^a\varphi(n)$. En effet, pour tout nombre premier p,

$$\varphi(p^{a+1}) = p^{a+1} - p^a = p^a(p-1) = p^a\varphi(p),$$

ce qui donne le résultat puisque φ est multiplicative.

3. La propriété de multiplicativité des degrés des extensions ([Pe], Corollaire 1.5 p. 65), donne

$$[\mathbb{Q}(z) : \mathbb{Q}] = [\mathbb{Q}(z) : \mathbb{Q}(\zeta)] \times [\mathbb{Q}(\zeta) : \mathbb{Q}].$$

Comme $[\mathbb{Q}(\zeta) : \mathbb{Q}] = \varphi(n)$, on obtient $[\mathbb{Q}(z) : \mathbb{Q}(\zeta)] = n^a$. Le polynôme minimal de z sur $\mathbb{Q}(\zeta)$ est donc de degré n^a. Or P_a est un polynôme annulateur de z, unitaire et de degré n^a, c'est donc le polynôme minimal de z sur $\mathbb{Q}(\zeta)$, et à ce titre, il est irréductible sur $\mathbb{Q}(\zeta)$. $\square$

Remarque : Le cas $n = 2$ redonne le résultat déja démontré au lemme 7.3.

B. Étude finale du spectre et conclusion

Lemme 7.4 *Soit P un polynôme à coefficients rationnels irréductible sur $\mathbb{Q}$. Si les racines complexes de P sont des racines de l'unité, alors P est un polynôme cyclotomique (à un inversible près).*

Preuve : Soit ζ une racine de P. C'est une racine n-ième de l'unité, donc une racine primitive mais pas forcément n-ième. Quitte à diviser par le coefficient du terme de plus haut degré, on peut supposer P unitaire. Il est de plus irréductible, c'est donc le polynôme minimal de ζ et coïncide donc avec un polynôme cyclotomique ([Pe], corrolaire 4.11 p. 83). □

Maintenant, tout est prêt pour la conclusion finale.

Théorème 7.8 (Spectre d'une matrice toute puissante)
Les valeurs propres d'une matrice TP$\mathbb{Q}$ *sont 0 ou 1.*

Preuve : Soit A une matrice TP$\mathbb{Q}$. On suppose qu'elle admet pour valeur propre une racine de l'unité ζ distincte de 1. On écrit alors $\chi_A = P^k \times Q$ avec P irréductible sur $\mathbb{Q}$ ayant ζ pour racine et P premier avec Q. En utilisant une base adaptée à cette décomposition via le lemme des noyaux, on obtient que A est semblable à une diagonale de 2 blocs qui sont eux même TP$\mathbb{Q}$ (on procède exactement comme dans la preuve de la proposition 7.2). Notons A' le premier bloc. Son polynôme caractéristique est justement P^k.
D'après la proposition 7.8, ses racines sont des racines de l'unité, donc d'après le lemme 7.4, P est un polynôme cyclotomique, et toutes ses racines sont dans $\mathbb{Q}(\zeta)$.
Ainsi $\chi_{A'}$ est scindé sur $\mathbb{Q}(\zeta)$, par dévissage (proposition 7.2), il existe une matrice nilpotente N telle que le bloc $\zeta I_k + N$ soit tout puissant sur $\mathbb{Q}(\zeta)$.
Pour tout $n \in \mathbb{N}^*$, il existe une matrice B à coefficients dans $\mathbb{Q}(\zeta)$ telle que $B^n = \zeta I_k + N$, ce qui donne $(B^n - \zeta I_k)^k = 0$ puisque N est nilpotente.
Le polynôme minimal de B, noté π_B, divise donc $(X^n - \zeta)^k$.
Notons t l'ordre de ζ, i.e. ζ est une racine primitive t-ième de l'unité. Prenons n égal à t^a de sorte que t^a soit strictement supérieur à k la taille de la matrice B (et donc au degré de π_B). La proposition 7.9 nous dit que $X^{t^a} - \zeta$ est irréductible sur $\mathbb{Q}(\zeta)$, comme $\deg \pi_B < \deg(X^{n^a} - \zeta)$, nécessairement, $\pi_B = 1$, ce qui est impossible. □

Ce théorème nous montre qu'une matrice TP$\mathbb{Q}$ a son polynôme caractéristique scindé sur $\mathbb{Q}$, on est donc toujours dans le cas du théorème 7.6. On a donc terminé, et démontré que les matrices toutes puissantes sur $\mathbb{Q}$ sont :

- les matrices unipotentes (cas inversible).

- les matrices semblables sur $\mathbb{Q}$ à une matrice bloc du type $\operatorname{diag}(I_r + N, 0)$ où N est une matrice rationnelle nilpotente et r le rang de la matrice.

Il est facile de voir que cette condition équivaut à dire qu'une matrice A de $M_p(\mathbb{Q})$ est TP$\mathbb{Q}$ si, et seulement si, $A(A - I_p)^p = 0$. On peut donc énoncer :

Théorème 7.9 (Description des matrices TP$\mathbb{Q}$**)** *Une matrice A de $M_p(\mathbb{Q})$ est toute puissante sur $\mathbb{Q}$ si, et seulement si, $A(A - I_p)^p = 0$.*

7.6.4 Annexe : nombres tout puissants sur un corps de nombre

Nous allons donc démontrer le théorème 7.7 de Perrin.
$\mathbb{K}$ désigne une extension finie de $\mathbb{Q}$. On note $\mathbb{A}$ l'anneau des entiers de $\mathbb{K}$, c'est-à-dire les éléments de $\mathbb{K}$ racines d'un polynôme à coefficients entiers et on note $\mathbb{A}^*$ le groupe de ses éléments inversibles (**les unités**). On sait que $\mathbb{A}$ est un anneau de Dedekind ([S], Ch. III, §. 4, Th. 1). L'idée est de copier la preuve utilisée pour $\mathbb{Q}$ (théorème 7.1), en remplaçant la décomposition unique en produit de nombres premiers par la décomposition en produit d'idéaux premiers. On commence par montrer le lemme suivant.

Lemme 7.5 *Soit $a \in \mathbb{K}$, $a \neq 0$ un nombre tout puissant sur $\mathbb{K}$. Alors a est dans $\mathbb{A}^*$, de plus ses racines n-ièmes sont aussi dans $\mathbb{A}^*$.*

Preuve : Pour tout $n \in \mathbb{N}^*$, il existe $b \in \mathbb{K}$ tel que $a = b^n$. On considère les idéaux fractionnaires non nuls (a) et (b). Par exemple,

$$(a) = \{ax |\, x \in \mathbb{A}\}.$$

Ils admettent des décompositions uniques en produits d'idéaux premiers ([S], Ch. III, §4, Th. 3) :

$$(a) = \prod_{P \in \mathcal{P}} P^{v_p(a)} \quad \text{et} \quad (b) = \prod_{P \in \mathcal{P}} P^{v_p(b)}$$

où $\mathcal{P}$ est l'ensemble des idéaux premiers non nuls de A, et où les $v_p(a)$ sont des entiers relatifs presque tous nuls.
On a $a = b^n$, donc $(a) = (b)^n$, l'unicité de la décomposition montre qu'on a pour tout P, $v_P(a) = nv_P(b)$. Cela montre que $v_P(a)$ est divisible par n et ce pour tout n, donc $v_P(a) = 0$ pour tout P, donc $(a) = (1)$ et donc que $a \in \mathbb{A}^*$. On obtient en plus que $(b) = 1$ et donc $b \in \mathbb{A}^*$. Les racines n-ièmes de a sont donc aussi des unités. Cela sera utile pour la suite. □

Remarque : Si N est une norme sur $\mathbb{K}$ ([S], Ch. II, §. 6), et que a est tout puissant, on a $N(a) = N(b)^n$ (la norme est multiplicative), et comme N est à valeurs dans $\mathbb{Q}$, cela implique que $N(a) = 1$ d'après le théorème 7.1.
Mais cela ne suffit pas à assurer que a est dans $\mathbb{A}^*$, penser à $c = \frac{3+4i}{5}$ dans $\mathbb{Q}(i)$. On a $N(a + ib) = a^2 + b^2$ donc $N(c) = 1$, pourtant c n'est pas entier sur $\mathbb{Q}(i)$ puisque les entiers de $\mathbb{Q}(i)$ sont les éléments de $\mathbb{Z}[i]$ ([S], Ch. II, §. 5, Th.1).

Pour finir, il reste le cas des unités. Dans $\mathbb{Q}$ il n'y avait que 1 et -1, ici les choses sont plus ardues :

Lemme 7.6 *Le seul élément de $\mathbb{A}^*$ qui est tout puissant sur $\mathbb{K}$ est 1.*

Preuve : Le théorème des unités de Dirichlet ([S], Ch. IV, §. 4, Th.1) précise la structure de $\mathbb{A}^*$. C'est un groupe multiplicatif abélien de type fini, précisément isomorphe à un groupe additif de la forme $\mathbb{Z}^k \times G$ où G est un groupe fini cyclique. Soit a dans ce groupe, on l'écrit additivement :

$$a = (a_1, ..., a_k, g)$$

avec $a_i \in \mathbb{Z}$ et $g \in G = \mathbb{Z}/r\mathbb{Z}$. Si a est tout puissant sur $\mathbb{K}$, d'après le lemme précédent, il est aussi tout puissant sur $\mathbb{A}^*$, ce qui se traduit en termes additifs par $a = nb$. On voit que ceci n'est possible que si les a_i et g sont tous nuls (pour g on prend $n = r$). On a donc $a = 0$ (au sens additif) donc $a = 1$ (au sens multiplicatif). □

7.7 Exercices corrigés

7.7.1 Diagonalisation de l'exponentielle d'une matrice

Exercice 7.1

1. Soit $A \in M_n(\mathbb{C})$. Donner la décomposition de Dunford de $\exp A$.

2. Montrer que $\exp A$ *est diagonalisable si et seulement si A est diagonalisable.*

3. Montrer que $\exp A = I_n$ *si et seulement si A est diagonalisable et* $\operatorname{spec} A \subset 2i\pi\mathbb{Z}$.

4. Juste pour rire sur la fin : donner la décomposition de Dunford de

$$M = \begin{pmatrix} -1 & 3 \\ 0 & 5 \end{pmatrix}.$$

Commentaires : Cet exercice permet de manipuler la décomposition de Dunford, il précise en plus le défaut d'injectivité de l'exponentielle.

Corrigé : 1. Soit $A = D + N$ la décomposition de Dunford de A. Comme D et N commutent, on a

$$\exp A = \exp D \exp N = \exp(D)\,(I_n + N') = \exp D + \exp(D)\,N'$$

avec

$$N' = N + \frac{N^2}{2!} + \cdots + \frac{N^{n-1}}{(n-1)!}.$$

Remarquons déjà puisque D et N sont des polynômes en A, que N' étant un polynôme en N est aussi un polynôme en A et que $\exp(D)$ étant un polynôme en D est aussi un polynôme en A.
La matrice $\exp D$ est diagonalisable car D l'est (si $D = P\,\mathrm{diag}(d_1, \ldots, d_n)P^{-1}$ alors $\exp(D) = P\exp(\mathrm{diag}(d_1, \ldots, d_n))P^{-1} = P\,\mathrm{diag}(\exp(d_1), \ldots, \exp(d_n))P^{-1}$).
La matrice N' est nilpotente comme somme finie de matrices nilpotentes qui commutent. De plus, N' commute avec $\exp D$ car ce sont des polynômes en A, donc $\exp(D)N'$ est nilpotente (en effet $(\exp(D)N')^n = \exp(D)^n N'^n = 0$ puisque $N'^n = 0$).
Enfin, les matrices $\exp D$ et $\exp(D)N'$ commutent aussi car polynômes en A, donc par unicité de la décomposition de Dunford, on conclut :

> La décomposition de Dunford de $\exp A$ est $\exp A = \exp D + \exp(D)N'$ avec $N' = \exp N - I_n$.

2. Si la matrice A est diagonalisable, on a déjà vu que $\exp A$ l'est aussi. Supposons que $\exp A$ est diagonalisable. Alors sa partie nilpotente $\exp(D)N'$ est nulle, donc $N' = \exp N - I_n = 0$, puisque $\exp D$ est inversible.
On peut alors conclure directement que $N = 0$ grâce à la bijection nilpotent unipotent (proposition 7.3) et donc que A est bien diagonalisable.
Sinon, on peut le faire à la main... Si N est non nulle d'indice de nilpotence r (avec $2 \leq r \leq n$), alors

$$N' = N + \frac{N^2}{2!} + \cdots + \frac{N^{r-1}}{(r-1)!} = 0,$$

donc X^r le polynôme minimal de N, divise $X + \frac{X^2}{2!} + \cdots + \frac{X^{r-1}}{(r-1)!}$, ce qui est impossible pour des raisons de degrés.

3. Le deuxième sens est facile, A est semblable à $\mathrm{diag}(a_1, \ldots, a_n)$ avec les a_i dans $2i\pi\mathbb{Z}$. La matrice $\exp A$ est donc semblable à $\mathrm{diag}(e^{a_1}, \ldots, e^{a_n}) = I_n$, donc $\exp A = I_n$.
Supposons que $\exp A = I_n$. Alors, d'après la question précédente, A est diagonalisable, donc semblable à $\mathrm{diag}(a_1, \ldots, a_n)$. Dans ce cas $\exp A$ est semblable à $\mathrm{diag}(e^{a_1}, \ldots, e^{a_n}) = I_n$. Ceci implique que pour tout $i \in \{1, \ldots, n\}$, $e^{a_i} = 1$ donc $a_i \in 2i\pi\mathbb{Z}$.

4. C'est facile. On écrit $M = \underbrace{\begin{pmatrix} -1 & 0 \\ 0 & 5 \end{pmatrix}}_{D} + \underbrace{\begin{pmatrix} 0 & 3 \\ 0 & 0 \end{pmatrix}}_{N}$.

Les matrices D et N sont respectivement diagonalisable et nilpotente. La décomposition de Dunford est le couple (D, N) à condition que D et N commutent ce qui est... **faux** ! En fait M a ses deux valeurs propres distinctes et

à ce titre elle est diagonalisable. Et puisque la matrice nulle est nilpotente et commute avec M, la décomposition de Dunford est en fait $M = M + 0$.

7.7.2 Racines carrées de matrices

Exercice 7.2 *Soit A une matrice de $M_n(\mathbb{R})$. On dit qu'une matrice R de $M_n(\mathbb{R})$ est une racine carrée de A si $A = R^2$. On note $\mathrm{Rac}\,A$ l'ensemble des racines carrées de A. Le but de cet exercice est de déterminer $\mathrm{Rac}\,A$ dans divers cas. On pourra dénombrer Rac A s'il est fini, sinon remarquer qu'il est constitué de classes de similitude.*

1. Déterminer $\mathrm{Rac}\,A$ dans le cas où A admet n valeurs propres réelles distinctes. Traiter l'exemple $A = \begin{pmatrix} 11 & -5 & 5 \\ -5 & 3 & -3 \\ 5 & -3 & 3 \end{pmatrix}$.

2. Déterminer $\mathrm{Rac}\,I_n$.

3. Déterminer $\mathrm{Rac}\,0$.

4. Déterminer $\mathrm{Rac}\,(-I_n)$.

5. Expliquer comment on peut déterminer $\mathrm{Rac}\,A$ dans le cas où A est diagonalisable.

6. Un cas non diagonalisable sur $\mathbb{R}$: déterminer $\mathrm{Rac}\,A$ dans le cas où $A = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ avec $b \neq 0$.

7. Étude topologique : on munit $M_n(\mathbb{R})$ d'une norme.

(a) Montrer que $\mathrm{Rac}\,A$ est une partie fermée d'intérieur vide de $M_n(\mathbb{R})$.

(b) $\mathrm{Rac}\,I_n$ est-elle une partie bornée ?

(c) Application : montrer que pour $n \geq 2$, il n'existe aucune norme $|| \ ||$ surmultiplicative sur $GL_n(\mathbb{R})$, c'est-à-dire telle que pour tous A et B dans $GL_n(\mathbb{R})$, $||AB|| \geq ||A|| ||B||$.

Commentaires : nous avons vu qu'une matrice inversible est TP$\mathbb{R}$ si et seulement si elle admet une racine carrée. On peut légitimement se demander s'il y a plusieurs racines carrées... On se propose de déterminer toutes les racines carrées de matrices diagonalisables. On verra qu'il peut y en avoir aucune, un nombre fini ou une infinité. Cet exercice est un bon entraînement aux diverses techniques de réduction, diagonalisation simultanée, utilisation de polynômes annulateurs...
L'exercice se termine par une petite étude topologique de Rac A. On y prouve une petite application qui nous dit qu'il n'existe pas de norme "surmultiplicative" sur $GL_n(\mathbb{R})$.

Corrigé : 1. A est diagonalisable, $A = PDP^{-1}$ avec D diagonale. Si $R^2 = A$, R commute avec A, donc laisse stable les **droites** propres de A (car A est à spectre simple), R est donc aussi diagonalisable dans la base d'espaces propres de A, d'où $R = PSP^{-1}$ avec S diagonale.
On a alors $S^2 = D$, ce qui donne $s_i^2 = d_i$, où (s_i) et (d_i) sont les coefficients diagonaux de S et D.
S'il existe un $d_i < 0$, alors il n'y a pas de racine carrée. Sinon, $s_i = \pm\sqrt{d_i}$ pour tout i.
Réciproquement les matrices $P\,\mathrm{diag}(\pm\sqrt{d_1}, \ldots, \pm\sqrt{d_n})P^{-1}$ sont bien des racines carrées de A.

Conclusion :

- si A admet une valeur propre strictement négative, A n'admet pas de racines carrées (réelles).

- si toutes les valeurs propres de A sont strictement positives, A admet 2^n racines carrées.

- si 0 est valeur propre de A et que ses autres valeurs propres sont positives, alors A admet 2^{n-1} racines carrées.

Dans l'exemple à traiter, on trouve $\chi_A(X) = -X(X-1)(X-16)$ et 4 racines carrées possibles :

$$\mathrm{Rac}\,A = \left\{ P \begin{pmatrix} 0 & 0 & 0 \\ 0 & \varepsilon_1 & 0 \\ 0 & 0 & 4\varepsilon_2 \end{pmatrix} P^{-1}, \quad \varepsilon_1, \varepsilon_2 \in \{\pm 1\} \right\}$$

avec $P = \begin{pmatrix} 0 & -1 & -2 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}$.

Remarque : attention, une matrice qui commute avec une matrice diagonalisable n'est pas forcément diagonalisable. En effet, toute matrice commute avec I_n, pourtant il y a des matrices non diagonalisables.

2. Si $R^2 = I_n$, R annule le polynôme $X^2 - 1$ scindé à racines simples. R est donc diagonalisable, semblable à $\mathrm{diag}(\varepsilon_1, \ldots, \varepsilon_n)$ avec les ε_i valant ± 1. Réciproquement, si $P \in GL_n(\mathbb{R})$, alors $R = P\,\mathrm{diag}(\varepsilon_1, \ldots, \varepsilon_n)P^{-1}$ vérifie bien $R^2 = I_n$. Donc

$$\mathrm{Rac}\,(I_n) = \{P\,\mathrm{diag}(\varepsilon_1, \ldots, \varepsilon_n)P^{-1}, \quad P \in GL_n(\mathbb{R})\}.$$

C'est la réunion de $n+1$ classes de similitude (cela correspond au nombre de 1 présents sur la diagonale de $\mathrm{diag}(\varepsilon_1, \ldots, \varepsilon_n)$).

3. Si $R^2 = 0$, R est nilpotente, donc semblable à une diagonale de blocs de Jordan J_k où

$$J_k = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix} \in M_k(\mathbb{R})$$

si $k \in \mathbb{N}^*$, et $J_1 = (0)$. Nécessairement la taille des blocs est inférieure ou égale à 2, sinon $R^2 \neq 0$. R est donc semblable à une matrice

$$R_k = \mathrm{diag}(\underbrace{J_2, \ldots, J_2}_{k \text{ fois}}, 0, \ldots, 0).$$

Pour des raisons de taille, il y a au maximum $\mathrm{E}(n/2)$ blocs J_2 dans R_k, donc $0 \leq k \leq \mathrm{E}(n/2)$. Réciproquement, toute matrice semblable à R_k vérifie bien $R_k^2 = 0$.
On remarque que rg $R_k = k$, donc si $k \neq k'$, R_k et $R_{k'}$ ne sont pas conjuguées. $\mathrm{Rac}\,0$ est donc constitué de $\mathrm{E}(n/2) + 1$ classes de similitude.

4. Commençons par une remarque : si $n = 2$, on interprète $-I_n$ comme la matrice de rotation d'angle π, on se dit alors que les matrices de rotation d'angle $\pi/2$ vont jouer un rôle.

• Si $R^2 = -I_n$, alors $(\det R)^2 = (-1)^n$, donc nécessairement n est pair.

• R annule le polynôme $X^2 + 1$ scindé à racines simples sur $\mathbb{C}$. La matrice R est donc diagonalisable, semblable sur $\mathbb{C}$ à $\mathrm{diag}(\varepsilon_1 i, ..., \varepsilon_n i)$ avec les ε_i dans $\{\pm 1\}$. Comme R est réelle, ses valeurs propres complexes sont conjuguées. Parmi les ε_i, il y en a donc autant qui prennent la valeur 1 que la valeur -1. Quitte à conjuguer par une matrice de permutation, on a donc

$$R \sim \mathrm{diag}(i, -i, ..., i, -i).$$

• Maintenant

$$R_{\pi/2} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \sim \mathrm{diag}(i, -i).$$

puisque $R_{\pi/2}$ a pour polynôme caractéristique $X^2 + 1 = (X - i)(X + i)$. Par produit de blocs, et par transitivité de la relation de similitude, on en déduit que $R \sim \mathrm{diag}(R_{\pi/2}, \ldots, R_{\pi/2})$.
Attention, pour le moment les 2 matrices sont semblables sur $\mathbb{C}$. Mais deux matrices réelles semblables sur $\mathbb{C}$, le sont aussi sur $\mathbb{R}$ ([2]).

[2] Rappelons la démonstration : si A et B de $M_n(\mathbb{R})$ sont semblables sur $\mathbb{C}$, il existe une matrice inversible P a priori à coefficients complexes telle que $AP = PB$. Cette matrice P

• Réciproquement, si n pair et $P \in GL_n(\mathbb{R})$, $R = P\,\mathrm{diag}(R_{\pi/2}, \dots, R_{\pi/2})P^{-1}$ vérifie bien $R^2 = -I_n$.

Faisons le bilan :

- Si n est impair, $\mathrm{Rac}\,(-I_n)$ est vide.
- Si n est pair, $\mathrm{Rac}\,(-I_n) = \{P\,\mathrm{diag}(R_{\pi/2}, \dots, R_{\pi/2})P^{-1},\ P \in GL_n(\mathbb{R})\}$ et l'on obtient une seule classe de similitude.

5. Soit R une racine carrée de la matrice A diagonalisable. L'idée est de se ramener aux cas précédents par diagonalisation simultanée possible grâce à la commutation.
Notons $\lambda_1, \dots, \lambda_k$ les valeurs propres de A et $p_1, \dots, p_k$ leur multiplicité respective. Notons u et v les endomorphismes de $\mathbb{R}^n$ canoniquement associés à A et R. Puisque u et v commutent, v laisse stable les sous-espaces propres de u. Ainsi dans une base de vecteurs propres de u (elle existe puisque u diagonalisable), la matrice de u est diagonale et celle de v est diagonale par blocs. Voici la traduction matricielle : si P désigne la matrice de passage de la base canonique de $\mathbb{R}^n$ à la base de vecteurs propres, on a

$$A = P\,\mathrm{diag}(\lambda_1 I_{p_1}, \dots, \lambda_k I_{p_k})P^{-1} \text{ et } R = P\,\mathrm{diag}(R_1, \dots, R_k)P^{-1}$$

avec $R_i \in M_{p_i}(\mathbb{K})$.
Comme $A = R^2$, par produit des blocs, on tire que pour tout $i \in \{1, \dots, k\}$ $\lambda_i I_{p_i} = {R_i}^2$. Les matrices R_i sont donc des racines carrées des matrices $\lambda_i I_{p_i}$.

- Si $\lambda_i = 0$, on est ramené à $\mathrm{Rac}\,0$
- Si $\lambda_i > 0$, on est ramené à $\mathrm{Rac}\,I_{p_i}$ car $I_{p_i} = \left(\dfrac{R_i}{\sqrt{\lambda_i}}\right)^2$.
- Si $\lambda_i < 0$, on est ramené à $\mathrm{Rac}\,(-I_{p_i})$ car $-I_{p_i} = \left(\dfrac{R_i}{\sqrt{-\lambda_i}}\right)^2$.

6. Si on pose $z = a + ib = re^{i\theta}$, A représente dans la base canonique la matrice de la similitude directe de rapport r et d'angle θ : $A = rR_\theta$ avec $R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.
On trouve alors immédiatement que les deux similitudes de rapport $\pm\sqrt{r}$ et d'angle $\theta/2$ fourniront deux racines carrées de A :

$$R_1 = \sqrt{r}R_{\theta/2} \quad \text{et} \quad R_2 = -\sqrt{r}R_{\theta/2}.$$

s'écrit $P = P_1 + iP_2$ avec P_1 et P_2 dans $M_n(\mathbb{R})$. Puisque A et B sont réelles, par identification des parties réelles et imaginaires, on a $AP_1 = P_1B$ et $AP_2 = P_2B$.

L'application polynomiale $\phi : t \mapsto \det(P_1 + tP_2)$ de $\mathbb{C}$ dans $\mathbb{C}$ est non nulle car $\phi(i) \neq 0$ donc admet un nombre fini de racines. En particulier, elle ne s'annule pas sur $\mathbb{R}$ tout entier, d'où l'existence d'un réel a tel que la matrice $Q = P_1 + aP_2$ soit dans $GL_n(\mathbb{R})$. On conclut alors facilement que $A = QBQ^{-1}$.

Reste à vérifier que ce sont les seules.
Il est facile de vérifier que A admet pour valeurs propres z et $\overline{z}$. Comme $b \neq 0$, cela fournit 2 valeurs propres distinctes. A est donc diagonalisable et à spectre simple sur $\mathbb{C}$. D'après la question 1, on obtient donc 4 racines carrées complexes de A semblables à :

$$\operatorname{diag}(\varepsilon_1 \sqrt{r} e^{\frac{i\theta}{2}}, \varepsilon_2 \sqrt{r} e^{\frac{-i\theta}{2}}) \quad \text{avec } \varepsilon_1, \varepsilon_2 \in \{\pm 1\}.$$

Reste à chercher lesquelles sont réelles parmi les 4. Si $\varepsilon_1 \neq \varepsilon_2$, les valeurs propres de la racine carrée ne sont plus conjuguées car de la forme s et $-\overline{s}$ avec $s \neq 0$, ce qui est impossible pour une matrice réelle.
A admet donc au maximum 2 racines carrées réelles, mais comme on a déjà vu que R_1 et R_2 conviennent, ce sont les seules. En conclusion :

$$\operatorname{Rac}(rR_\theta) = \{\pm \sqrt{r} R_{\theta/2}\}.$$

7.(a) Rac A est un ensemble algébrique, c'est-à-dire une intersection de zéros de fonctions polynomiales non nulles. A ce titre, il est fermé et d'intérieur vide (c'est une petite partie au sens topologique de Baire).
Ce résultat est très utile, une preuve détaillée figure dans l'épreuve d'algèbre du concours CCP section MP de 2005. Donnons néanmoins les grandes lignes de la preuve. Un ensemble algébrique est fermé comme intersections de fermés (qui sont des images réciproques du fermé $\{0\}$ par une application continue). Un ensemble de zéros d'une fonction polynomiale P non nulle est d'intérieur vide. En effet s'il admet un point intérieur, P s'annule sur une boule ouverte qui est égale (en choisissant la norme infinie) à un produit cartésien d'intervalles ouverts, ce qui implique alors que le polynôme est nul. On conclut ensuite puisqu'une intersection de parties d'intérieur vide est encore d'intérieur vide.

7.(b) Montrons que pour $n \geq 2$, Rac I_n n'est pas borné. On choisit la norme infinie sur $M_n(\mathbb{R})$, qui vaut le max des valeurs absolues des coefficients. Pour tout entier p non nul,

$$R_p = \begin{pmatrix} -1 & p \\ 0 & 1 \end{pmatrix}$$

est une racine carrée de I_2, de norme égale à p.
Si $n > 2$, la matrice $\operatorname{diag}(R_p, 1, \ldots, 1)$ est encore une racine carrée de I_n de norme p, ce qui prouve que pour $n \geq 2$, Rac I_n n'est pas borné.

7.(c) Soit $n \geq 2$. Remarquons déjà qu'il n'existe pas de norme surmultiplicative sur $M_n(\mathbb{R})$, car il y a des diviseurs de zéros : si A et B sont non nulles et telles que $AB = 0$, on a $||A||\,||B|| \leq ||AB|| = 0$, donc $||A||\,||B|| = 0$ donc

$A = 0$ ou $B = 0$ qui est faux[3].
Supposons qu'il existe $|| \; ||$ surmultiplicative sur $GL_n(\mathbb{R})$ (où il n'existe plus de diviseurs de zéros). Soit R une matrice de $\mathrm{Rac}\, I_n$. Elle est inversible car $\det R = \pm 1$. On a alors

$$||I_n|| = ||R^2|| \geq ||R||^2$$

et donc $||R|| \leq \sqrt{||I_n||}$, ce qui est absurde puisque $\mathrm{Rac}\, I_n$ n'est pas borné (on utilise le fait que, dans un espace vectoriel de dimension finie, toutes les normes sont équivalentes).

7.7.3 Raffinement de la surjectivité de l'exponentielle

Exercice 7.3 *Cet exercice propose une jolie preuve topologique du lemme 7.2 : pour toute matrice A de $M_n(\mathbb{C})$, il existe un polynôme P de $\mathbb{C}[X]$ tel que $A = \exp(P(A))$.*

1. En dimension 1 :

(a) Montrer que $H = \exp(\mathbb{C})$ est un sous-groupe ouvert de $\mathbb{C}^*$.

(b) Montrer que H est aussi fermé dans $\mathbb{C}^*$, conclure que $\exp : \mathbb{C} \to \mathbb{C}^*$ est surjective.

2. En dimension quelconque : Soit $A \in M_n(\mathbb{C})$. On veut montrer qu'il existe $P \in \mathbb{C}[X]$ tel que $A = \exp(P(A))$.

(a) L'application exp est-elle un morphisme de groupe additif $M_n(\mathbb{C})$ sur le groupe multiplicatif $GL_n(\mathbb{C})$?

(b) On note $\mathbb{C}[A]^* = \mathbb{C}[A] \cap GL_n(\mathbb{C})$. Justifier que $\exp : \mathbb{C}[A] \to \mathbb{C}[A]^*$ est bien définie et que $\mathbb{C}[A]^*$ est un ouvert de $\mathbb{C}[A]$.

(c) Montrer que exp est de classe C^1 sur $M_n(\mathbb{C})$.

(d) Montrer que $H = \exp(\mathbb{C}[A])$ est un sous-groupe ouvert de $\mathbb{C}[A]^*$, puis conclure.

Commentaires : nous reprenons ici la jolie preuve de Pommelet [Po] de la surjectivité de exp de $\mathbb{C}$ sur $\mathbb{C}^*$. Si on essaye de la généraliser à $GL_n(\mathbb{C})$, ce qui bloque c'est le défaut de commutativité, pour pallier à ce problème, l'idée est de restreindre la source à un ensemble de matrices qui commutent, on choisit alors naturellement $\mathbb{C}[A]$.

[3] Cette remarque aboutit en fait à une autre preuve : supposons donc qu'il existe une norme surmultiplicative sur $GL_n(\mathbb{R})$. Soit A une matrice inversible fixée et B une matrice non nulle telle que $AB = 0$. Par densité de $GL_n(\mathbb{R})$ dans $M_n(\mathbb{R})$, il existe une suite de matrices inversibles (B_k) qui converge vers B.

Pour tout entier naturel k, on a $||A|| \, ||B_k|| - ||AB_k|| \leq 0$, d'où par passage à la limite (licite puisque l'application $M \mapsto ||A|| \, ||M|| - ||AM||$ est continue), on a $||A|| \, ||B|| \leq ||AB|| = 0$ d'où $||A|| \, ||B|| = 0$, et donc $A = 0$ ou $B = 0$, ce qui est faux.

Cet exercice utilise le théorème d'inversion locale, et une technique de connexité dans les groupes topologiques[4]. Plus précisément, on verra que si H, sous-groupe d'un groupe topologique G, admet un point intérieur, alors H est ouvert dans G mais il est alors aussi fermé dans G, donc si G est connexe...

Corrigé : 1.(a) Pour tout $a, b \in \mathbb{C}$, $\exp(a+b) = \exp(a)\exp(b)$. La fonction $\exp : \mathbb{C} \to \mathbb{C}^*$ est donc un morphisme de groupes, par suite $\exp(\mathbb{C})$ est un sous-groupe de $\mathbb{C}^*$.
Montrons que 1 est intérieur à H. La fonction exp est holomorphe de dérivée elle même, elle est donc de classe C^1 si on la regarde comme une fonction de deux variables réelles, sa différentielle en 0 est la multiplication par $\exp(0) = 1$, c'est donc l'identité sur $\mathbb{R}^2$. D'après le théorème d'inversion locale, exp est donc localement inversible, il existe un voisinage ouvert $\mathcal{V}$ de 1 dont tous les éléments sont des exponentielles, donc $1 \in \mathcal{V} \subset H$, ce qui montre que 1 est intérieur à H.
Si $a \in H$, $a\mathcal{V}$ est un voisinage de a, ouvert puisque image de $\mathcal{V}$ par l'homéomorphisme $h \mapsto ah$ (sa réciproque est $h \mapsto a^{-1}h$), et inclus dans H car H est stable par multiplication. H est donc ouvert.

1.(b) On partitionne G en ses classes modulo H ($x \sim y \Leftrightarrow xy^{-1} \in H$). Chacune de ses classes bH est ouverte comme image de l'ouvert H par $h \mapsto bh$. Le complémentaire de H dans $\mathbb{C}^*$ est donc la réunion des ouverts bH avec $b \notin H$, c'est donc une partie ouverte de $\mathbb{C}^*$, ce qui donne H fermé dans $\mathbb{C}^*$.
Concluons : H est à la fois fermé et ouvert dans $\mathbb{C}^*$, qui est connexe, et H n'est pas vide, donc $H = \mathbb{C}^*$.

2.(a) Dès que $n \geq 2$, comme deux matrices ne commutent pas forcément, on n'a plus $\exp(A+B) = \exp(A)\exp(B)$, l'application $\exp : M_n(\mathbb{C}) \to GL_n(\mathbb{C})$ n'est donc plus un morphisme ! On ne peut donc pas généraliser ainsi la preuve. Il nous faut de la commutativité. On pense alors aux algèbres de polynômes...

2.(b) Si $M = P(A)$ où P est un polynôme,

$$\exp(M) = \lim_{n\to+\infty} \sum_{k=0}^{n} \frac{(P(A))^n}{n!}.$$

Comme $\mathbb{C}[A]$ est une $\mathbb{C}$-algèbre de dimension finie[5] ($\mathbb{C}[A]$ est un sous-espace vectoriel de $M_n(\mathbb{C})$ de dimension finie sur $\mathbb{C}$), c'est donc une partie fermée de $M_n(\mathbb{C})$. Par suite $\exp(M)$ est un polynôme en A, puisque limite d'une

[4] Un groupe topologique est un groupe muni d'une topologie séparée pour laquelle la multiplication et l'inverse sont continues. Cette définition n'est pas utile pour l'exercice.

[5] En fait la dimension de $\mathbb{C}[A]$ vaut le degré de π_A, polynôme minimal de A.

suite de polynômes en A. On sait en plus que $\exp(M)$ est inversible, donc $\exp : \mathbb{C}[A] \to \mathbb{C}[A]^*$ est bien définie.
On a $\mathbb{C}[A]^* = \{M \in \mathbb{C}[A],\ \det M \neq 0\}$, donc c'est une partie ouverte de $\mathbb{C}[A]$ comme image réciproque de l'ouvert $\mathbb{C}^*$ par l'application continue det (car polynomiale).

2.(c) Une méthode consiste à utiliser le théorème de dérivation terme à terme d'une série de fonctions. On pose pour $k \in \mathbb{N}$,

$$f_k : M \mapsto \frac{M^k}{k!}.$$

Les fonctions f_k sont de classe C^1 sur $M_n(\mathbb{C})$ (car les coefficients de la matrice $M^k/k!$ sont des polynômes en les coefficients de M), la série $\sum_{k\geq 0} f_k$ converge simplement sur $M_n(\mathbb{C})$. Il suffit donc de prouver que montrer que la série $\sum_{k\geq 0} \mathrm{d}\, f_k$ converge uniformément sur les compacts de $M_n(\mathbb{C})$.
Pour cela, on choisit sur $M_n(\mathbb{C})$ une norme $||\ ||$ sous-multiplicative, c'est-à-dire telle que $||AB|| \leq ||A||\,||B||$.
Il nous faut calculer pour tout entier $k \geq 1$, la différentielle de la fonction f_k, donc de l'application $M \mapsto M^k$. Pour la différentielle en X, on isole les termes linéaires en H obtenus en développant $(X+H)^k$. Les autres termes (en nombre fini) ont au moins deux fois le "facteur" H, ce sont donc des $O(||H||^2)$.

$$\begin{aligned}(X+H)^k &= (X+H)\cdots(X+H)\\ &= X^k + HX^{k-1} + XHX^{k-2} + \cdots + X^{k-1}H + O(||H||^2).\end{aligned}$$

On en déduit pour $k \geq 1$ que la différentielle de f_k en X, notée $\mathrm{d}f_k$, est l'application linéaire de $M_n(\mathbb{C})$ dans $M_n(\mathbb{C})$ définie par

$$\mathrm{d}f_k(X).H = \frac{1}{k!}(HX^{k-1} + XHX^{k-2} + \cdots + X^{k-1}H).$$

On note $\mathcal{L}$ l'espace vectoriel des applications linéaires de $M_n(\mathbb{C})$ dans $M_n(\mathbb{C})$, que l'on munit de la norme $\mathcal{N}$ subordonnée à $||\ ||$. Il s'agit de montrer que $\sum_k \mathrm{d}f_k$ converge uniformément sur les compacts de $M_n(\mathbb{C})$. Pour tout H appartenant à $M_n(\mathbb{C})$, on a

$$||\mathrm{d}f_k(X).H|| \leq \frac{1}{k!}k||X||^{k-1}||H||,$$

donc

$$\mathcal{N}\left(\mathrm{d}f_k(X)\right) \leq \frac{||X||^{k-1}}{(k-1)!}.$$

La série $\sum_{n\geq 0} \mathrm{d}f_k$ est donc uniformément convergente sur les boules

$$\{X \in M_n(\mathbb{C}),\ ||X|| \leq R\}$$

de $M_n(\mathbb{C})$, la série $\sum_{k\geq 0} f_k$ converge simplement vers exp, et l'on en déduit que exp est de classe C^1 sur ces boules, donc sur $M_n(\mathbb{C})$.

Remarques : α) Vous pouvez regarder Rouvière ([Ro], exercice 38), pour une preuve du théorème de dérivation terme à terme des séries de fonctions.

β) On peut montrer que exp est de classe C^∞, on pourra consulter Lafontaine ([Laf], p 36) pour une preuve différente à l'aide d'analyse complexe.

γ) Attention, on ne peut justifier que exp est de classe C^1 par l'argument $\exp A$ est un polynôme en A. En effet, les coefficients du polynôme dépendent de A... Avec ce même raisonnement, on comettrait l'erreur de dire que l'application qui à une matrice associe son polynôme minimal est continue... Ceci est faux, la suite

$$A_n = \begin{pmatrix} 0 & 1/n \\ 0 & 0 \end{pmatrix}$$

tend vers la matrice nulle. Pour tout n, le polynôme minimal de A_n est X^2, il ne peut donc tendre vers X le polynôme minimal de la matrice nulle.

2.(d) A partir de là, il n'y a plus qu'à imiter la preuve de la question 1. Comme tous les polynômes en A, commutent entre eux, exp : $\mathbb{C}[A] \to \mathbb{C}[A]^*$ est un morphisme de groupes, et $H = \exp(\mathbb{C}[A])$ est un sous-groupe de $\mathbb{C}[A]^*$. La fonction exp est de classe C^1, sa différentielle en 0 est l'identité donc est inversible :

$$\exp(H) = I_n + H + O(||H||^2).$$

On en déduit l'existence d'un voisinage ouvert $\mathcal{V}$ tel que $I_n \in \mathcal{V} \subset H$. Pour tout $M \in H$, $M\mathcal{V}$ est un voisinage ouvert de M dans H, car $B \mapsto MB$ est un homéomorphisme sur $GL_n(\mathbb{C})$ (c'est une application linéaire donc continue car en dimension finie), d'inverse $B \mapsto M^{-1}B$).
Le groupe H est donc un sous-groupe ouvert de $\mathbb{C}[A]^*$, il est donc aussi fermé (même preuve qu'en 1.b).
Si M et N sont dans $\mathbb{C}[A]^*$, la fonction polynomiale non nulle de $\mathbb{C}$ dans $\mathbb{C}$

$$z \mapsto \det\left((1-z)M + zN\right)$$

n'admet qu'un nombre fini de zéros et ne s'annule ni en 0, ni en 1, donc il existe un chemin continue γ joignant 0 à 1 qui évite ces zéros dans $\mathbb{C}$. Alors l'arc paramétré

$$t \mapsto (1-\gamma(t))\, M + \gamma(t) N$$

joint M et N et est à valeurs dans $\mathbb{C}[A]^*$, ce qui prouve que $\mathbb{C}[A]^*$ est connexe par arcs.
Finalement la partie H est non vide, ouverte et fermée dans le connexe $\mathbb{C}[A]^*$, donc $H = \mathbb{C}[A]^*$, ce qui achève la preuve.

Remarque : $GL_n(\mathbb{C})$ est un groupe topologique, sa topologie provient de l'espace vectoriel normé $M_n(\mathbb{C})$, et les applications $(A, B) \mapsto AB$ et $A \mapsto A^{-1}$ sont continues (comme $A^{-1} = \frac{{}^t\text{com}A}{\det A}$, l'application inverse est une fraction rationnelle en les coefficients de A).

7.7.4 Matrices toutes puissantes sur $\mathbb{Z}$

Exercice 7.4 *Le but de cet exercice est de déterminer les matrices toutes puissantes sur $\mathbb{Z}$, à savoir les matrices A de $M_n(\mathbb{Z})$ telles que pour tout $k \in \mathbb{N}^*$, il existe une matrice B de $M_n(\mathbb{Z})$ telle que $A = B^k$.*
Si p est un nombre premier, on note $\overline{x}$ la classe d'un entier x modulo p, et π_p l'application de $M_n(\mathbb{Z})$ dans $M_n(\mathbb{Z}/p\mathbb{Z})$ qui à une matrice M de coefficients (m_{ij}) associe sa réduction modulo p, c'est-à-dire la matrice $\pi_p(M)$ dont les coefficients sont $(\overline{m_{ij}})$.

1. Démontrer que si $A \in M_n(\mathbb{Z})$ est toute puissante sur $\mathbb{Z}$, alors pour tout nombre premier p, la matrice $\pi_p(M)$ est toute puissante sur $\mathbb{Z}/p\mathbb{Z}$.

2. Démontrer que si une matrice M de $M_n(\mathbb{Z})$ est telle que pour tout nombre premier p, $\pi_p(M) = 0$, alors M est la matrice nulle.

3. Démontrer que les matrices de $M_n(\mathbb{Z})$ toutes puissantes sur $\mathbb{Z}$ sont les matrices de projecteurs (on pourra utiliser la détermination des matrices toutes puissantes sur les corps finis).

Commentaires : ce joli petit exercice est un exemple réussi de passage du local au global en arithmétique. On cherche tout d'abord des solutions modulo p (solutions "locales") puis on remonte aux solutions sur $\mathbb{Z}$ (solutions "globales"). On peut citer pour la culture le *principe de Hasse* : sous certaines hypothèses, une équation polynomiale $P(x) = 0$ a des solutions dans $\mathbb{Q}$ si et seulement si elle en a dans tous les complétés de $\mathbb{Q}$ (au sens métrique du terme), et les seules complétions de $\mathbb{Q}$ sont $\mathbb{R}$ et les corps p-adiques $\mathbb{Q}_p$ pour tout p premier.

Corrigé : 1. Soit p un nombre premier. Soit $M = (m_{ij})$ et $N = (n_{ij})$ deux matrices de $M_n(\mathbb{Z})$. Puisque la surjection canonique

$$\begin{array}{ccc} \mathbb{Z} & \to & \mathbb{Z}/p\mathbb{Z} \\ x & \mapsto & \overline{x} \end{array}$$

est un morphisme d'anneau, on a

$$\overline{m_{ij}+n_{ij}} = \overline{m_{ij}} + \overline{n_{ij}} \quad \text{et} \quad \overline{\sum_{k=1}^{n} m_{ik}n_{kj}} = \sum_{k=1}^{n} \overline{m_{ik}}\ \overline{n_{kj}}.$$

Ainsi $\pi_p(M+N) = \pi_p(M) + \pi_p(N)$ et $\pi_p(MN) = \pi_p(M)\pi_p(N)$ et donc l'application

$$\begin{array}{rccc} \pi_p: & M_n(\mathbb{Z}) & \to & M_n(\mathbb{Z}/p\mathbb{Z}) \\ & (m_{ij}) & \mapsto & (\overline{m_{ij}}) \end{array}$$

est un morphisme d'anneaux.
Soit A toute puissante sur $\mathbb{Z}$. Pour tout $k \in \mathbb{N}^*$, il existe une matrice B de $M_n(\mathbb{Z})$ telle que $A = B^k$. Alors en réduisant modulo p, il vient par le morphisme π_p, $\pi_p(A) = \pi_p(B)^k$, ainsi $\pi_p(A)$ est toute puissante sur $\mathbb{Z}/p\mathbb{Z}$.

2. Soit $M = (m_{ij})$ dans $M_n(\mathbb{Z})$ telle que pour tout nombre premier p, $\pi_p(M) = 0$. Alors les coefficients m_{ij} sont des entiers divisible par tout nombre premier p, ce qui n'est possible que si $m_{ij} = 0$, c'est-à-dire $M = 0$.

3. Soit A une matrice toute puissante sur $\mathbb{Z}$. Alors d'après la question 1, pour tout nombre premier p, la matrice $\pi_p(A)$ est toute puissante sur le **corps fini** $\mathbb{Z}/p\mathbb{Z}$. Mais on a démontré au théorème 7.5 que les matrices toutes puissantes sur un corps fini sont les projecteurs. Ainsi $\pi_p(A)$ est un projecteur et $\pi_p(A)^2 = \pi_p(A)$. D'où par morphisme $\pi_p(A^2 - A) = 0$ et ceci pour tout nombre premier p. D'après la dernière question, cela implique que $A^2 - A = 0$, et donc A est un projecteur.
Réciproquement si $A \in M_n(\mathbb{Z})$ est un projecteur, par récurrence immédiate pour tout $k \in \mathbb{N}^*$, on a $A = A^k$ donc A est toute puissante sur $\mathbb{Z}$.

Références :

[Laf] Lafontaine Jacques, Introduction aux variétés différentielles, PUG (1996).

[Mn] Mneimné Rached, Réduction des endomorphismes, Calvage et Mounet (2006).

[Pe] Perrin Daniel, Cours d'algèbre, Ellipses (1996).

[Po] Alain Pommellet, Cours d'analyse, agrégation de mathématiques, Ellipses (1998).

[Ro] François Rouvière, Petit guide de calcul différentiel à l'usage de la licence et de l'agrégation, Collection : enseignement des mathématiques, Cassini (1999).

[S] Samuel Pierre, Théorie algébrique des nombres, Hermann (1967).

[www] Coste Michel, http ://agreg-maths.univ-rennes1.fr/documentation/docs/Exponentielle.pdf.